

# What to do with the former Soviet Union

The world seems bent on helping the former Soviet Union through its present crisis, but is not sure what to do. There is a special need where science is concerned — and the international research community has special responsibilities.

Mr George Bush, the President of the United States, should not be surprised that even his friends are mad at his handling of last week's international conference in Washington on the provision of help for the former Soviet Union. For one thing, there was no meaningful agenda. For another, the potential beneficiaries of unprecedented compassion had not been asked. Finally, Bush himself diminished the proceedings by producing a cheque for \$600 million, as if to challenge others in attendance to get out their cheque-books. What might have been a serious occasion became a hurried charity auction.

The business of helping the former Soviet Union is at once more matter-of-fact and more important. Charitable considerations are irrelevant. The truth is that the West, by which is meant the world outside the former Soviet Union, needs an accommodation with Russia, which comprises three-quarters of the population of the former union (and, because of the land mass involved, an even larger proportion of its natural resources). By comparison with Russia, the recently liberated states of Central Europe are small potatoes; that does not mean that they are unimportant, merely that they are smaller. (One of the obvious dangers to be avoided in the present circumstances is that the continuing needs of these Central European states will be neglected: they should not be.) But the former Soviet Union was also a nuclear power, sufficiently so that for decades people have been kept too late at work at places such as the Pentagon. There needs also to be an understanding on that issue, and quickly.

Last week's proposal from the United States that up to 2,000 ex-Soviet military scientists may be harmlessly occupied for a few years in dismantling ex-Soviet nuclear bombs, for which the US Congress has already set aside \$500 million, is sensible. It should help to take the edge off a discarded community's frustrations in the next few years. But the recipe is not a foolproof way of preventing the spread of nuclear technology from the former Soviet Union. There will always be people willing to make bombs for, so to speak, the hell of it — and inducements large enough to tempt them. Only a strengthening of the physical controls on nuclear materials, backed up by the governments' determination to give proliferators a hard time, can accomplish that.

But what Russia needs in the next few months — there is no point in imagining a more distant future — is the mix of food and medicines called "humanitarian aid" last

week. Otherwise, people will needlessly starve or die from other causes. Economic aid is also needed, although not in the way that Mr Boris Yeltsin has been talking of: his "fund" to stabilize the ruble will be a black hole until he and his bankers stop printing money. Why not give the man an overdraft of \$10,000 million until April, on the understanding that he might then wash his hands of it, at a price (political or fiscal) then to be determined?

But economic aid as such will not help ex-Soviet science. Other claims on the attention of the Russian government may well take precedence, some say rightly. That is why it will eventually fall to the international research community to see to it that Russian science gets the help it needs. For the time being, there are some particular needs that deserve wide attention:

**First, unsurprisingly, journals, any journals; the recipients will usually ensure that copies are forwarded, if necessary, to more appropriate recipients. Just imagine what it must be like, to be seeking to throw back the frontiers of ignorance without knowing quite where they are!**

**Second, travel. People, and especially young people, should be going to conferences overseas, or visiting colleagues and competitors elsewhere. At present, such luxuries are beyond reach (see page 391). One solution is that the world's airlines with rights to fly to Russian and other ex-Soviet cities should offer spare seats on a standby basis, at no cost to those who fill them. A better way would be a computerized market in cost-free vacant airline seats. Any such system would cost the airlines next to nothing, but would earn endless if admittedly intangible goodwill.**

**The largely French-inspired plan to create a foundation within the former Soviet Union also deserves backing. The idea is to make research grants to able scientists. It is to be hoped that the scheme will not be confined to collaborative projects, which are often needlessly cumbersome and expensive. Some measure of personal support should be included in the grants. It is to be hoped that as many governments and private foundations as possible can be persuaded to join the scheme. What members of the research community tell government officials could be influential.**

There is also much that Russia and the other republics can do to help themselves. In Moscow, a new minister of science appears to have made an energetic beginning on



the restructuring of the research enterprise (see page 383), but there is still a long way to go. It is a pity that the reconstitution of the Russian Academy of Sciences in the past few weeks has not produced the tighter organization that circumstances require. The problem of the relationship between research and the universities — or the general lack thereof — also deserves attention. It is tempting to suppose that the middle of an economic crisis is not the time at which to embark on the long-term institution-building that will eventually be necessary. But this, surely, is just the time for planning the more creative relationship between the universities and research that is needed. Crises like that now in Russia are not just depressing times but are opportunities as well. □

## An obdurate recession

THE finance ministers of the seven richest nations of the world (called G7) seem to have held a meeting bordering on the complacent in New York last weekend. Their purpose was to “intensify their cooperative efforts to strengthen world economic growth”. But the finance ministers, who were accompanied by their central bankers, seem as puzzled as the next man or woman about the causes of an obdurate recession. And they seem to have found no recipe for restimulating growth, except the decision that each member of the group will now aim, in necessarily different ways, at “sustainable growth with price stability”. It is thus when petty criminals promise the magistrates who let them off lightly that they will do better in future.

As economic recessions go, the present (which set in a year ago) is not especially severe. Gross domestic product in some G7 countries has fallen by a few per cent, but in others (Germany and Japan for example) it is merely that the previous rate of growth has not been sustained. And whether the present downturn is essentially different from its predecessors remains a matter for conjecture. It is worrying that, both in the United States and Britain, signs of improvement have proved to herald a false spring. Last week's unemployment figures in the United States showed an unexpected increase when the government had been expecting (not to say hoping) that they would go the other way. None of this should surprise the finance ministers and central bankers of G7. It is not, after all, as if the world's financial system was in such good shape before the recession began. Among other things, it followed a long period in which bank and other financial institutions reported huge losses, first on loans to developing countries, then nearer home. Seemingly paper transactions, such happenings take money out of the system — witness the now-greater caution of most banks. For a time, impoverishment may have been concealed by the funds injected into the system to offset the supposed deflationary consequences of the stock-market crash of October 1987. When governments then took up the battle against inflation by means of high interest rates, two deflationary influences

were brought together.

Since then, there has been a steady stream of bank and company failures adding to the spread of a general sense of impoverishment. When people see assets disappearing, it is only natural that they should spend less freely, and save instead. What with mounting fears of unemployment, and often the reality of it, there is every chance that people's attitudes towards the use of money have shifted substantially, perhaps permanently, in spendthrift countries such as Britain and the United States. Even if the banks would let them borrow, they are disinclined to spend. That is a prospect G7 should have reckoned with last weekend. So is the inevitable diversion of funds to Eastern Europe and further east.

So what, if anything, could get growth moving again? The answer is buried in the G7 statement put out from New York — a successful conclusion of the negotiations of the new deal under the General Agreement on Tariffs and Trade (GATT). The consequences could be apparent by the summer. By securing the more economic distribution of goods and (on this occasion) some services, many new suppliers would enjoy a greater sense of being rich, as would the customers for their goods, who would be paying lower prices for them. Of all the remedies for recession considered by G7, only a deal on trade can break the logjam. The group should have said so, and then have pinned the blame for impending failure on the European Communities' backwardness on agriculture. □

## Setting priorities

THROUGHOUT the world where government budgets are constrained by recessionary pinch, it is necessary to establish priorities in science as in other areas of public life. It is not enough to say that science is important; it is necessary to acknowledge that some research may be of greater importance (or value) than others.

In the United States, where president George Bush has just released his budget plans for the 1992 fiscal year, some scientific groups are trying to lead the way to a structure for setting priorities within science. For this, consensus within the affected community is required. The National Research Council has just taken a (modest) step with a report on space that says why it is important for scientists to set priorities (if they do not someone else will) but leaves actual priority-setting to a future committee.

The US National Institutes of Health is also getting into the priority setting business with a ‘strategic plan’ (whose details will be announced next week) that calls upon biomedical researchers to agree that funds for some areas should grow at a rate that is higher than others. The pretence that this does not happen already should give way to a more realistic view of the world of science funding. Setting priorities may be difficult (if for no other reason than offending one's colleagues) but in the interests of science itself it must be done. Efforts to set priorities should be supported, or at least viewed with an open mind. □



# Roche cuts controversial PCR fees, testing limits

- Researchers fought high prices, restrictions
- Firm promises reforms, may increase profits

## Cold Spring Harbor, New York

UNDER pressure from researchers and diagnostic companies, Swiss-owned pharmaceutical company Hoffmann-La Roche last week agreed to lower its prices on its polymerase chain reaction (PCR) technology and lift restrictions on its use.

The move should free the technology, which is a cornerstone of the biotechnology industry, for widespread use in testing for paternity and genetic and infectious diseases. Although PCR, a method for amplifying DNA, has been freely used in research laboratories since the mid-1980s, researchers say that its use in diagnostics has been hamstrung by restrictive and expensive licensing demands from Roche and PCR's previous owner, Cetus Corp.

Roche now charges small diagnostic laboratories \$24 or 15 per cent of the test price (whichever is more) for each test, something that has contributed to a typical test price of over \$100, says Thomas Reed, chairman of Vivigen, a small New Mexico-based diagnostic company. Because of the flat-fee minimum rate, he says, lowering prices only increases the proportion of license costs.

At a meeting at Cold Spring Harbor laboratory last October, Roche officials came under fire for its practices from laboratory director James Watson. Watson also warned of congressional hearings on the situation and at least one other researcher has suggested that PCR is so valuable a "national resource" that the government should take the patent by eminent domain, rather than allow one company to control its availability.

Under such criticism, Roche has decided to release its hold on the market. It intends to lower its fees to "less than 10 per cent" for non-profit diagnostic centres such as health clinics and academic hospitals and laboratories, and slightly more than that for large diagnostic companies, which now make up more than 90 per cent of the market, says Thomas White, vice president for research and development at Roche Molecular Systems, a new Roche subsidiary set up to market PCR. There will be no minimum flat fee. Current restriction on the use of the technology for paternity testing and testing for infectious diseases, such as AIDS, and Lyme disease, will be also lifted, he says. Roche says it intends to announce the full details of the new policy within the next several weeks.

Company officials and outside analysts say that the move is not entirely altruistic. Although Roche said its licensing restrictions were based on concerns about testing standards at diagnostic laboratories and legal complications, analysts say that the lure of a monopoly may have played a larger role in its policy. Roche has its own paternity and infectious-disease testing laboratories that, of course, were not subject to most of the restrictions.

But as the potential applications of PCR have grown, the economic picture has changed. Now the company is expected to make far more money on PCR by taking a small cut of a large, unencumbered market than by trying to monopolize a smaller market with exclusive licenses and high fees. Roche paid \$300 million for PCR last year, an amount that many analysts felt was too much for a patent that will expire in slightly more than a decade. "It's an awfully expensive toy and the only way they'll get their money back is by licensing it freely," says Kary Mullis, who invented the technology when he was at Cetus and is now an independent biotechnology consultant.

Furthermore, by unleashing PCR in what is soon expected to be a diagnostic market worth over \$100 million, Roche may be able to pre-empt competition from Abbott Laboratories, which is developing an alternative gene amplification technique known as ligase chain reaction (LCR). Although LCR may turn out to be better than PCR for certain disease testing applications, its principal selling point has been that it would free diagnostic laboratories from Roche's PCR monopoly.

Roche's policy shift is expected to have its greatest impact on diagnostic testing for genetic diseases. Although that market is now relatively small and focused on prenatal testing for conditions such as Down syndrome, new genetic mapping and sequencing technology is speeding the rate at which new disease genes are being located and hastening the day when widespread genetic screening is commonplace.

The US Human Genome project has started a trial of mass screening for cystic fibrosis (CF) using a PCR-based method, but high and restrictive licensing fees have slowed acceptance of such testing. Combined with other fees — the discoverers of the CF gene intend to charge a six per cent

royalty for each CF test — total license fees for combined disease testing could conceivably amount to "\$70 of a \$75 test" says Arthur Beaudet, a molecular geneticist who runs a diagnostic laboratory at the Baylor College of Medicine. "Things are just out of control." Vivigen, for example, is currently losing money on its tests. Reed hopes that the Roche decision will mean the beginning of more rational policies across the board. "I've been hanging on, waiting for an announcement like this," he says.

Perkin-Elmer Corp., which makes PCR equipment and holds the rights to the use of PCR for forensic tests such as DNA "fingerprinting", said it would also revamp its licensing policies in the coming weeks. In a statement released last week, the company promised a "clearer and better-refined licensing program" on the forensic tests, but revealed no details.

**Christopher Anderson**

## JAPANESE SCIENCE

### Frontiers win

#### Tokyo

US GOVERNMENT officials announced last week in Tokyo that the United States plans to contribute \$3.5 million to Japan's Human Frontier Science Program in fiscal 1993. The injection of US funds will give a significant political as well as financial boost to the programme which supports international research on the brain and molecular biology.

The announcement of US support, which was expected although a clear dollar amount had not been specified until now (*Nature* 353, 685; 24 October, 1991), was made at a government officials meeting on Frontiers on 21 and 22 January attended by representatives of the seven western summit nations (Japan, United States, Canada, the United Kingdom, France, Germany and Italy), the European Community, and Switzerland.

Until now nearly all of the annual Frontier budget of approximately \$30 million per year has been provided by Japan with only small contributions from some of the other participant nations. The US contribution, which will come from the National Institutes of Health, NASA, the Department of Energy, and the National Science Foundation, will make the United States the second largest contributor to the programme after Japan.

On hearing the news, Japanese government officials involved in the programme expressed relief that the US contribution does not seem to be linked to Japanese participation in the US Superconducting Super Collider (SSC) project, as they once thought it might be. Japan has yet to make a decision on the SSC.

**David Swinbanks**



# Shakeup for Royal Observatories

## London

A SCIENCE and Engineering Research Council (SERC) plan to bring Britain's two separate Royal Observatories under the control of a single, powerful director is dividing UK astronomers. Many staff at the Royal Greenwich Observatory (RGO) in Cambridge and the Royal Observatory, Edinburgh (ROE) believe the new organization will destroy the institutes' traditional autonomy, bringing them more strongly under the influence of SERC's headquarters at Swindon. Given the frequently aired misgivings of SERC's chairman Sir Mark Richmond about the high proportion of the council's budget spent on 'big science', they fear the worst.

Some university astronomers share their colleagues' concern, arguing that the loss of the scientific leadership traditionally given by the two Royal Observatory directors could undermine morale, seriously damaging British astronomy. But others believe changes are long overdue, and say that the new central control could lead to better access to Britain's telescopes for university astronomers. The debate will intensify over the next few months, as SERC searches among its own employees for a suitable candidate for the new job.

The Royal Observatories pre-date the modern organization of UK science, and were once synonymous with British astronomy. (RGO dates from the seventeenth century, ROE from the eighteenth.) But the growth of a university-based British astronomical community, and the observatories' takeover by SERC in the 1960s, forced changes. Over the past decade, SERC's policy has been to reduce the observatories' status as premier research centres (the number of staff employed at the two sites has fallen from 341 to 245 since 1985). Today, their main role is to run large telescopes for use by the wider UK astronomical community. For RGO, this means running Britain's optical telescopes, at La Palma, in the Canary Islands, while ROE is responsible for a series of infrared telescopes in Hawaii.

But according to SERC's deputy chairman, Tony Hughes, the separation between the two observatories' roles is now inappropriate. The next major project in British astronomy — the Gemini collaboration with the United States and Canada (see *Nature* 353, 688; 1991) — is to build two 8-metre telescopes that will observe in both the visible and infrared spectra, cutting across the traditional demarcation between the two Royal Observatories. And

some astronomers say privately that the natural spirit of competition between the two Royal Observatories has not always served British astronomy well.

Some Royal Observatory astronomers view SERC's plans in a much more sinister light, however. As SERC employees, they are unwilling to speak publicly, fearing retribution. But amid dark mutterings about SERC's "management by terror", the fear is that the new director may not fight strongly for the observatories' interests against a SERC leadership seen by

run temporarily by Paul Murdin, formerly Boksenberg's deputy at RGO.

Longair agrees with SERC that the work of the two observatories needs pulling together. But he believes that SERC should be careful to appoint a director who will have the support of the staff. There is a problem with morale, he says, and "if these people get discouraged, they may go elsewhere".

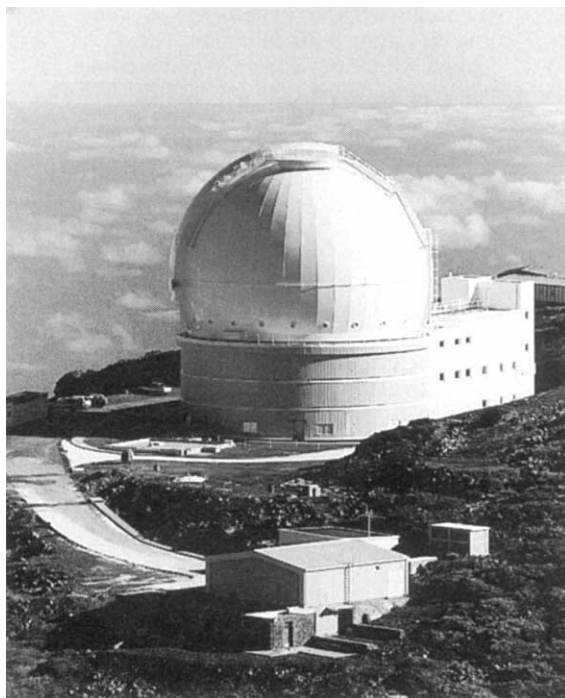
Other university astronomers are worried about the possibility of an administrator, rather than a scientist, having such a powerful position in British astronomy. "That's not the way to run any scientific enterprise," contends Sir Martin Rees, from the Institute of Astronomy at Cambridge, pointing out that top-class laboratories such as CERN and Fermilab, though equally complex to administer, are headed by internationally respected scientists. He is concerned that the downgrading of the RGO and ROE directors' posts will erode the scientific leadership at the two observatories.

But university astronomers are divided over the issue. If the observatories are to co-operate more closely, "what they need is a manager," says Ken Pounds, from the University of Leicester, who is president of the Royal Astronomical Society. The best solution would be to find an astronomer who is also a good manager. But "are there such people?" he asks. Pounds (himself an X-ray astronomer, so not a user of the Royal Observatories' telescopes) views the observatories' present organization as an historical accident, which does not necessarily ensure the best access to the island telescopes for university astron-

omers. "Remember that [SERC's] establishments should be there to help the universities carry out research."

The Astronomer Royal, Arnold Wolfendale from the University of Durham, says he fully supports the SERC plan — he heads the council's Astronomy and Planetary Science Board. He even believes that the new organization could encourage SERC to spend more on astronomy, if UK astronomers now begin to "pull in the same direction". Given the depth of resentment among many Royal Observatory staff, however, that may be a tall order.

**Peter Aldhous**



The 4.2-metre William Herschel Telescope at La Palma, in the Canary Islands run by RGO.

many astronomers as hostile to the subject.

SERC wants to give the job to one of its existing employees. And as the new Royal Observatory head's job is a managerial position, an administrator, rather than a respected astronomer, is thought likely to get the job (although Hughes stresses that no decision has yet been made). The directors of RGO and ROE "have provided very much of a buffer between the bureaucracy in Swindon and the astronomical community," says one RGO astronomer. "Now that buffer is being removed." RGO staff also fear that their current director, Alec Boksenberg (widely respected for his role in developing the Hubble Space Telescope's Faint Object Camera), may have no place in the new set-up. For its part, ROE has had no permanent director since Malcolm Longair left for the University of Cambridge in 1990, and is being

## CORRECTION

In the News story "'IQ pills' land in trouble" (*Nature* 355, 285; 23 January 1992), the authors of the article in *The Lancet* should have been given as David Benton and Gwilym Roberts.

# Leukaemia linked to radiation

## London

THE UK government's radiation watchdog, the National Radiological Protection Board (NRPB), has produced the first convincing evidence linking low-level exposure to radiation at work with an increased incidence of fatal leukaemia. The study, published in the *British Medical Journal* (304, 220; 25 January 1992), suggests that the risk of contracting leukaemia following exposure to low levels of radiation over many years is about twice that calculated from previous studies. Nevertheless, the total number of excess deaths is thought to be relatively low.

The results have led to calls for the tightening of British and international guidelines on the maximum annual radiation dose for nuclear workers. The International Commission on Radiological Protection (ICRP)'s recommendation that workers should receive no more than 20 milliSieverts (mSv) a year, averaged over a five-year period (see *Nature* 348, 274; 1990), is based on studies of the Japanese atom bomb survivors — who received radiation in a single, intense exposure. But NRPB director Roger Clarke, presenting the new results at a press conference last week, said that the uncertainty surrounding the NRPB's new risk estimate is too great to sanction immediate changes to the British or international guidelines.

Although the link between leukaemia and occupational exposure to radiation has been explored before, with inconclusive results, the new NRPB study is unique in its sheer scale. In 1976, the board set up the National Registry of Radiation Workers (see *Nature* 255, 517; 1975), which documents the radiation doses received by some 95,000 past and present workers in the British nuclear industry. The new NRPB study links these dose data with cause of death, for 6,660 workers who had died by the end of 1988.

NRPB epidemiologists found a significant positive correlation between radiation dose and death from leukaemia (excluding chronic lymphatic leukaemia, which is not thought to be inducible by radiation). They also found a similar — but not statistically significant — trend for fatal solid tumours. Barbara MacGibbon, NRPB's medical director, last week said that it will take several years to determine whether this trend is real: the longer lag time between exposure to radiation and the development of solid cancers means that the number of workers who have so far died is too small a sample for a conclusive analysis.

The radiation-leukaemia link will strengthen the cases of tens of leukaemia patients now trying to obtain compensation from their employers in the nuclear industry. But Clarke last week said that the

NRPB's results should be put in perspective. Despite the association between radiation dose and leukaemia within the nuclear industry, nuclear workers are less likely to die from the disease than the general British population. This is due to the industry's tendency to employ people in good health, a 'healthy worker effect' that was repeated for most of the causes of death examined by the NRPB group. MacGibbon estimated that the number of excess deaths from leukaemia caused by exposure to radiation in the British nuclear industry since the late 1940s is less than ten (from a total of 47 workers in the NRPB's national registry who died of leukaemia).

Nevertheless, Patrick Green, radiation campaigner with the UK environmentalist group Friends of the Earth, last week called

for the NRPB's recommended upper limit for radiation exposure at work to be cut from 15 mSv to 10 mSv a year. He is also worried about a forthcoming European Communities' (EC) directive, which is expected to adopt the ICRP's 20 mSv-a-year average as a legally binding limit for the 12 EC states. But Clarke, who is an expert adviser to Euratom, the EC agency responsible for the directive, rejected this idea. Until more workers in the national registry die of leukaemia, allowing NRPB to calculate with greater confidence the risk of developing fatal leukaemia after radiation exposure, Clarke argued that it would be inappropriate to reject the current ICRP guidelines. He agreed that the NRPB leukaemia risk estimate, if confirmed, would increase the pressure for a 10 mSv-a-year limit, but he thought it more likely that the risk would be "confirmed at the levels we're working with at the moment".

**Peter Aldhous**

## DRUG DEVELOPMENT

# Orphan drug windfalls?

## Washington

ALTHOUGH most orphan drugs have annual sales of less than \$3-5 million, some members of Congress are concerned that a handful of companies are exploiting the exclusive marketing provisions provided by the Orphan Drug Act and charging patients exorbitant prices. As a result, companies such as Amgen, Inc., Genentech, Inc., Genzyme Corporation, Eli Lilly and Company and Fujisawa, Inc. are making windfall profits from cumulative sales in the United States that top \$200 million. Bills have been introduced in both the House and Senate that, while retaining the key incentives of the act, would open up the market to competition for commercially-viable orphan drugs.

Primarily through the guarantee of a government-granted monopoly for a seven-year period, the Orphan Drug Act sets out to foster the development of drugs for diseases that affect fewer than 200,000 people and are likely to be of limited commercial interest. In the nine years since its enactment, the US Food and Drug Administration (FDA) has approved 60 orphan drugs for marketing. By contrast, only ten drugs were developed for the treatment of rare diseases in the 10 years before its enactment.

Although Congress is focusing its attention on four orphan drugs that are racking up block-buster sales (human growth hormone, Epogen, aerosol pentamidine, and Ceredase), at least another eight orphan drugs already on the market or nearing approval have the potential to become block-buster products.

Under the proposed amendments to the act, the orphan status of a drug would

be revoked and competitors allowed to enter the market when cumulative net sales of the drug exceeded a \$200-million cap. Similar legislation introduced in the House sets the cap at \$150 million.

Legislative uncertainty that has surrounded the Orphan Drug Act over the past 3 or 4 years is already having an effect on the number of new orphan drugs being developed. An FDA official says there has been a 10 per cent drop in the number of orphan drug applications in 1991 compared to the previous year.

At a hearing last week before a congressional subcommittee, chaired by Senator Howard Metzenbaum (Democrat, Ohio), Genzyme and Genentech were asked to justify the high prices they are charging for their orphan drugs. Amgen, Eli Lilly and Fujisawa, though invited to testify, declined to do so.

Ceredase, manufactured by Genzyme for the treatment of Gaucher's disease, costs patients an average of \$250,000 in the first year of treatment. Since its approval by the FDA last April, total sales for Ceredase are estimated to be \$120 million. Although Genzyme executives refused to disclose to the subcommittee details of research and development costs for the drug, Henri Termeer, chairman and chief executive officer of Genzyme, admitted that Metzenbaum's \$50-million estimate was "approximately right". Termeer blames high manufacturing and other incidental costs for the high price.

Even if a version of the bill is passed by Congress, it will still face the daunting hurdle of winning sufficient votes to overcome a threatened presidential veto.

**Diane Gershon**



# A gene library that goes "moo"

## Washington

THEY are hardly rare and certainly far from exotic, but barnyard animals may soon be the next "endangered species" to get an international conservation programme.

The United Nation's Food and Agricultural Organization (FAO) is preparing to launch a five-year, \$15-million Global Animal Genetic Resources Programme that aims to rescue such unlikely animals as the Shiwal cow of Pakistan, the extraordinarily ugly Taihu pig of China and the stringy Fayoumi chicken of Egypt. Each is threatened not by predators or lost ecosystems, but by established Western breeds such as the Holstein cow and the factory chicken. Hoping to increase their output, farmers in developing countries are switching to high-productivity Western animals, abandoning native breeds that have been adapting to the local conditions for decades. FAO worries that if the local breeds are allowed to disappear, their disease-resistance traits and ability to withstand harsh conditions may be lost, too.

Although Western farm animals generally produce more milk, eggs or offspring than the hardened stock of Africa, Asia and Latin America, experience has shown that there is nothing like an epidemic or a heat wave to remind local farmers why the old breeds have been around for so long. The Chinese Taihu pig may not grow to the size of a typical Western sow, but it can live on cabbage if need be.

This week, E. Patrick Cunningham, director of the FAO's animal production and health division, met officials of the World Bank in Washington, hoping to have the animal genetic resource programme included in the World Bank's \$1,300 million Global Environment Facility programme, which was established last year to promote conservation and protection efforts around the world. Although there are many current efforts to protect the genetic resource of agricultural plants — a more threatened group than the animals because of the ease in which seeds can be exchanged — the FAO project is one of the first to target livestock.

The first priority, says Cunningham, will be to find out what now exists. FAO is building a database of livestock breeds, including such data as the kind of farming conditions in which they are used and their birthing intervals. This, it turns out, is not always as easy as it sounds. "There are countries where all they can say is that 'there are sheep out there,'" he says.

Once the FAO researchers have the list in hand, they plan to identify the breeds most in danger of extinction. The Kerry cow of Ireland, for instance, is now down to only 200 animals. Since the turn of the century, half of Europe's domestic breeds

have been lost, and a third of the remaining 770 breeds are in danger of disappearing within the next two decades. In parts of the developing world, the situation may be as bad or worse.

The tropics are likely to be the most critical area. Because that is where human populations are growing fastest, tropical farmers are under the greatest pressure to improve production. Switching to a high-output Holstein may help for the moment, but FOA is worried about the possibility of disease or drought; pathogens are being spawned as fast as people in that part of the world. Researchers say that it is only a matter of time before one of them emerges to wipe out the Western stock. By then, the disease-resistant traits of some once-local breed may have already been lost.

FAO proposes to use about a dozen species — including the Barbados Blackbelly sheep, the N'Dama cattle of West

Africa and the South American Criollo cattle — as trial projects in field preservation. Other endangered breeds would be added as they are identified.

DNA 'fingerprinting' of rare breeds can indicate which are distinct enough from more common stock to merit preservation. And as geneticists continue to identify disease genes, it may soon be possible to find out which breeds have the most valuable resistance traits. "We have no illusion," says Cunningham, "that this kind of money will allow us to do even a pale shadow of the genome project [the international effort to map human DNA]. But we want to steer and stimulate the people who are already doing this" to bring some attention to agricultural animals. Several groups — including those at the Indian Veterinary Research Institute, the University of Brisbane and Texas A&M University — are already investigating livestock genetic resources. FAO hopes to get more researchers to join them down on the farm.

**Christopher Anderson**

## EC BIOTECHNOLOGY POLICY

# Progress on animal patents

## London

A DRAFT European Communities (EC) directive that would endorse the patenting of transgenic animals has finally cleared its most important hurdle. After nearly three years of inactivity, the European Parliament's legal affairs committee has decided that the 'inventors' of transgenic animals should be able to file patents in the EC. European Commission biotechnology officials believe that the directive, first proposed in 1988, will now be adopted by ministers from the EC member states before the end of 1993.

The directive has been controversial since its inception, attacked by the European Green movement and animal rights groups, who argue that many transgenic animals, particularly those produced for commercial gain, are likely to suffer. Patent Concern, a coalition formed specifically to oppose the directive, promises to continue the fight. But Willi Rothley, the German Social Democrat Member of the European Parliament who wrote the legal committee's report on the directive, believes the full parliament — which votes in February — will back his committee's position "by a large majority".

Ironically, the protracted delays in the parliament may have worked to the directive's advantage: two years ago, opinion about the wisdom of allowing animal patents was sharply divided. But now that the European Patent Office (EPO) has decided to grant a patent for Harvard University's transgenic 'oncomouse' (see *Nature* 353, 589; 1991), the political climate has changed. An EC decision to ban

animal patents would now place the eight EC member states that have also signed the EPO's charter, the European Patent Convention, in a tricky legal position.

Although the parliament's legal committee rejected Patent Concern's call to ban animal patents, two amendments drafted by Rothley would place some restrictions on patenting. Chimaeras (animals produced by combining cells from two or more species into a single embryo) and genetic modifications that cause "unnecessary suffering...or physical harm" should not be patentable, the legal committee decided.

European Commission patent experts are worried that Rothley's wording would exclude Harvard's oncomouse, which is genetically engineered to develop tumours. The oncomouse clearly suffers physical harm, but is a useful tool in cancer research. Commission officials want to avoid inconsistency between the EC directive and EPO's decision, and say the amendments will have to be changed, after the directive leaves the parliament in February. But the changes are not expected to cause serious problems: the next time the parliament votes on the directive, it can register its disapproval only by rejecting the directive outright. This is considered unlikely, as the directive also contains a number of measures to improve patent protection for the European biotechnology industry that command widespread support. And in any case, Rothley says that his amendments are not designed to make the oncomouse patent illegal within the EC.

**Peter Aldhous**



# New broom sweeps cleaner

## Moscow

Mr Boris G. Saltykov, Minister of Science, University and Technology Policy in the new Russian government, is energetic, thoughtful and imaginative. But circumstances require that he must also be a realist. A balanced budget requires that the money-printing presses should be stopped and that there will be hard times for many people and institutions when that happens. One prospect is that jobs will be lost. Saltykov says he expects the number of people working for the academic sector, chiefly that of the Russian Academy of Sciences, to have declined by between 20 and 30 per cent by the end of 1993. "It is obligatory", he said in his ministry office last Friday.

The government, Saltykov said, does not have the money to support "the whole old structure of research". People repeatedly visit him asking for extra funds for particular projects, but "I have no additional money". Instead, "I have promised to help with getting help" from outside.

Saltykov's analysis of the present situation is simple: there is a struggle, he says, between different parts of society. "The simple people on the streets" have no particular wish to support science, "especially after Chernobyl". And, in a democracy, public opinion counts for something.

With some passion, the minister nevertheless insists that it is in the interests of "world science" that help should materialize. It is not just that, otherwise, many people from defence research will end up assisting with the proliferation of defence technology but that the world will be the poorer if Russian science goes to the wall.

Saltykov hopes for much from the foundation for assisting ex-Soviet science advocated by M. François Mitterrand, the President of France, among others, partly at the urging of Dr Carlo Rubbia, the director-general of the European high-energy physics laboratory (CERN) in Geneva. The hope is that decisions about the foundation, especially about its scope, will be reached during February.

Meanwhile, he seems to be doing what he can to put existing institutions on a clearer footing. Thus Russian membership of CERN will continue, thanks to a deal with Rubbia allowing the contribution of SFr600,000 to be postponed for a year: in return, Saltykov has agreed to continue to support the home bases, chiefly at Serpukhov and Dubna, of Russians collaborating on experiments at CERN, who may be as many as 20 at any one time. Even more generously, the nine Russians at DESY, the electron accelerator laboratory in Hamburg, will be maintained from German funds during the year ahead.

But high-energy physics in general does not have a bright future. The electron

collider being built at Serpukhov (and already two years late) will not be completed unless the economic climate improves. There is, for the time being, no chance that the 40-km-long colliding machine, based on two opposed linear accelerators, will ever be built.

Dubna is a different case (see page 391). The objective here is to breathe "new life" into a laboratory too much concerned with high-energy physics and too little with general and applicable physics. Among other things, the minister has taken the unprecedented step of advertising for a new director on the open market.

Another reorganization that seems on paper already to be a success is that of the instrument technology institute at Leningrad, previously an awkward joint venture between the Academy of Sciences and a production ministry. The institute has now been founded as a joint-stock company, 40 per cent of which has been sold to IBM and Deutsche Bank. The institute's strong suit is the design of sophisticated instruments but, to cut its commercial teeth, it is talking with suppliers in the Far East about the manufacture of personal computers.

That project raises a complaint from Saltykov — about the committee called CoCom, based in Paris, with continuing responsibility for licensing the export of strategic materials to eastern Europe and to the ex-Soviet Union. He says he can understand why it came into being, but not why it continues to superintend such a wide range of devices — custom-built chips for instruments, for example. He suspects it is another example of people hanging on to obsolete jobs.

On the academic front, change will be harder to bring about. Acknowledging that the Soviet Academy had become "a kind of science ministry", Saltykov was implicitly acknowledging that his own existence has robbed the academy of part of its function. One gathers that the minister (who was an academy servant in charge of a science policy study before becoming a politician) will let the academy work out its own solutions (or salvation) within the confines of the budget.

Earlier last week, Academician Yu. Ossipov, the newly elected president of the Russian Academy, had said that his first task is to "preserve the academy", but even that does not imply no change. In reality, opinion within the academy is divided. Casual conversation uncovered two academicians holding that the academy should withdraw from much of the applied research for which it is now responsible and also form stronger bridges with the universities. (The second said he was the only academician to hold these radical views, but, told that there was at least one

other, instantly identified the first informant, suggesting that his claim to uniqueness may be almost correct.)

On links between research and the universities, there seems general agreement that change must be slow. Saltykov says that Russia has inherited roughly 600 of the 900 institutions of university rank in the old Soviet Union, of which a few, "maybe 20 or 30, are really excellent". The proportion, he says, is not very different from the proportion of research universities in the United States. He lists among the outstanding universities the Moscow State University, the Moscow Physical-Technical Institute (his own alma mater) and Leningrad University. Part of the obstacle to change is the fear among university teachers that they would be intellectually swamped if researchers also habitually taught. Saltykov would like to see some change, but perhaps only when other problems are less pressing.

The problem of defence research is one of these, and for good reasons. Defence research is estimated to employ 35 per cent of Russian scientists, and to have taken 80 per cent of research funding. It has traditionally been well paid, but now there is no point in turning out "new tanks just to destroy them". One step along the road to pacification has been to convert the Kurchatov Institute, a kind of Los Alamos in suburban Moscow, into a "national centre", but there is plainly a long way further to go.

So with everything in Russia, Saltykov repeatedly returned to the dominance of the economic problem, which he often called a "crisis". If that is not solved, no amount of toiling in the vineyards of a ministry will count for much. Saltykov knows that, but too many of his constituents are unpersuaded. **John Maddox**

## ENVIRONMENT

### UK sets 'early warning system'

#### London

NINE UK government agencies, led by the Natural Environment Research Council (NERC), are setting up a new 'early warning system' designed to detect environmental change. Starting with eight stations, the Environmental Change Network (ECN) aims to monitor a range of British terrestrial ecosystems, recording a suite of data from simple meteorological measurements to samples of soil invertebrates. The network is expected to help to define future UK government environmental policy, but more stations may have to be added to distinguish between local environmental fluctuations and national trends. Relatively few countries have set up similar monitoring networks. The new British effort joins similar monitoring schemes in China, Sweden and the United States. P.A.

# Freedom does not bring money

SIR — Freedom does not bring money. This truth is trivial, but we in post-Communist countries meet it every day, and in its most concentrated form in science. Only a few years ago, scientists from mid- and East-European countries were treated as animals in cages to be looked at and mildly fed. Many friendly scientists in the free world tried to help them by invitations to the other side of Iron Curtain, paying their expenses at symposia and congresses, by gifts of scientific books, arranging the exchange of scientific journals and so on. Then, the representatives of the Communist regimes tried to end these contacts, not only from fear of the infiltration of adverse ideology, but also because proud socialist science could not go begging.

Now we are free, which is a good feeling. But freedom has not brought money to our countries. Especially in science, we are now even more poverty-stricken than before. And this situation is not realized even by our friends in economically strong countries. "It is apparent to me that Dr X has not been able to keep up with developments in the subject over the last few years. I suggest that he should now, in the light of the great freedom that has come to Czechoslovakia, take advantage of the new situation and visit, particularly, the Agricultural University at Uppsala in Sweden. . . ." wrote a British friend in a review for the journal *Photosynthetica*, which I edit. But can Dr X really do that?

Let us omit the worst cases, such as the rest of Soviet Union or Romania, which still struggle both for real freedom and for bread. Take Czechoslovakia, generally understood abroad to be in a more or less acceptable economic situation. Parliamentary democracy functions, with all its advantages and shortcomings, and a clever policy keeps inflation within reasonable limits. But prices have increased by at least 50 per cent, while our (traditionally low) salaries have increased by only some 15 per cent. The government put an end to the black market in foreign currency by making the old black market exchange rate the official rate. Thus the life of foreign guests in Czechoslovakia is 2–3 times cheaper than before, but our visits to Western countries are 2–3 times more expensive.

The budgets of research institutes and universities are more or less the same as in Communist days, but the prices we pay for instruments, chemicals, literature and so on are much higher. This has already led, in the institutes of the Czechoslovak Academy of Sciences, to a 25 per cent reduction of staff, because

otherwise we would not even have heat and light in our laboratories. At Czechoslovak universities, there has been no such reduction of staff; indeed, five new universities were founded in 1991. Their situation may be even worse than ours. But can scientists pay for themselves when the institutions cannot?

Here are a few examples of our present situation. My salary is one of the best in our institute, but at the present rate of exchange (50 Kč's = £1) I get something like £100 per month after deducting tax (a postgraduate student gets £40 and a science beginner some £60). It is not difficult to calculate that my membership subscription to the American Society of Plant Physiologists, including the journals *Plant Physiology* and *The Plant Cell*, costs more than my monthly salary, a week's stay in the United Kingdom at a symposium with fees about three times my monthly salary, buying a PC half a year's salary and so on. Most of us are accustomed to having only a few pounds in our pockets, eating fast food and even sleeping in a laboratory when abroad, and we do not complain. Nevertheless, the travelling costs are possible only by Czechoslovak bus (Prague–Paris–Prague for a third of my monthly salary). It is much more expensive to use a train (the same trip second class without a sleeping berth costs more than a month's salary) or plane (more than two months' salary). So it is almost impossible to travel overseas.

So our newly acquired freedom is somewhat illusory and the wish of many scientist for better knowledge, contacts and common research projects still remains in the realm of dreams. I ask our old and new scientific friends to realize that freedom does not necessarily bring a horn of plenty and to support us in the same way — or better — than before. We have to return to the world science community as soon as possible, and we are in no better an economic situation than the developing countries you often help.

ZDENĚK ŠESTÁK

*Institute of Experimental Botany,  
Czechoslovak Academy of Sciences,  
160 00 Prague 6,  
Czechoslovakia*

SIR — We appeal on behalf of the recent Congress of the Soviet Astronomical Society, which was established in 1990 and which unites about 400 career astronomers. It is the only independent organization to represent the professional interests of about 3,000 Soviet astronomers. We wish to draw attention to the rapidly deteriorating situation of astro-

nomy in our country (by which we mean the whole territory of the former Soviet Union, including the now independent republics).

Astronomy is likely to become one of the first victims of the economic crisis, if only because it involves considerably fewer people than other sciences. Financial support for many scientific projects has almost ceased. Our small community may collapse in a few months without immediate support. We also have great difficulty in getting funds for subscriptions to astronomy periodicals.

Nevertheless, we have able theoreticians as well as observers, and there exist a number of unique instruments and small and medium-sized telescopes over a large range of longitude. Mutually profitable co-operative studies with astronomers elsewhere could be developed in the fields of theoretical astronomical and astrophysical research. Compilation of astronomical catalogues is a possibility, as well as different branches of observational works, including those that do not require sophisticated technology but need a long timespan, such as variable star and some extragalactic researches. Joint investigations and grants to help to overcome intellectual isolation from the scientific community may be the best way of helping us to survive for the next several critical years and to conserve our potential for the benefit of the world astronomical community.

On behalf of the Congress of the Soviet Astronomical Society, we hope to hear from institutions or individuals willing to help to organize forms of co-operation and support for Soviet astronomy and astronomers.

N. G. BOCHKAREV  
A. D. CHERNIN  
YU. N. EFREMOV  
N. N. SAMUS'  
A. V. ZASOV

*Soviet Astronomical Society,  
Sternberg Astronomical Institute,  
13 Universitetskij Prospekt,  
119899 Moscow, USSR  
(Fax: 095 939 0126)*

## Left out

SIR — Lists such as those of scientific centennials and half-centennials (J. L. Heilbron and W. F. Bynum, *Nature* 355, 11; 1992) are not intended to be exhaustive but I think the article is a little unfair to the memory of Henry Walter Bates, discoverer of batesian mimicry, who died in London on 16 February 1892.

RICARDO FERREIRA

*Department of Physiology,  
Michigan State University,  
East Lansing,  
Michigan 48824-1101, USA*



# Science against the odds in Poland

Poland's new science policymakers are struggling to build a genuinely competitive system of science funding. But political instability threatens the pace of reform, as economic slump decimates the science budget.

*Warsaw, Krakow & Gdansk*

TODAY, the triumphal entry of the erstwhile dissidents of the Solidarity movement into government, only two and a half years ago, seems a distant memory. In the 1989 elections, Solidarity swept the board, packing the Senat, the upper house of parliament, with its candidates, and winning all of the fairly-contested seats — but not enough for an overall majority — in the more-powerful lower house, or Sejm. But those days of euphoria and unity are far behind. The first post-Communist government, led by former Solidarity activist Tadeusz Majowiecki, lasted only until autumn 1990, when Majowiecki resigned as prime minister after his defeat by Lech Walesa in Poland's first ever presidential election. And although Poland held its first fully democratic parliamentary elections in October 1991, the prospects for strong government to tackle the country's severe economic problems are poor. More than 20 different parties now hold seats in the Sejm, with the largest single party — Majowiecki's centre-left Democratic Union — winning only 12 per cent of the seats. Not surprisingly, coalition building has been a fraught business, and although Jan Olszewski did win sufficient parliamentary support to become Poland's new prime minister in the week before Christmas (beating off the challenge of former businessman Jan Krzysztof Bielecki, leader of the previous temporary government appointed by Walesa), his centre-right coalition looks shaky.

With Poland's attempt to kick start a market economy after forty years of central planning running into problems — unemployment is growing and the annual rate of inflation has surged to over 50 per cent — all that seems certain now is that the long-suffering Polish people face further hardship. Poland's politicians now find themselves torn between the desire to protect the electorate from the worst ravages of unemployment and price increases, and the need to keep the national budget deficit within targets set by the International Monetary Fund. Some former Solidarity activists look

wistfully back to the unity of purpose of 1989, but the splintering of Solidarity (as a political movement, rather than a trade union) is perhaps a natural reaction to decades of imposed political conformity, and some Poles say that compromise and alliance-building run counter to the Polish national character.

So what are the prospects for science, in the midst of this political and economic turmoil? One positive factor is that the

government after the 1989 elections, the old mechanisms of science funding were quickly broken down. Under the Communists, state support for research came through the budgets of a series of government ministries — industry, health, environment, education and the like — and the Academy of Sciences, which funded research in its 80 institutes. But now, all Polish government research spending has been handed to a single body, the state Committee for Scientific Research (KBN).

Established in early 1991, KBN is an unusual experiment in democracy. Although its chairman is appointed by the Sejm, most of its members were elected in a ballot of nearly 30,000 active PhD-level researchers. KBN is supported by two commissions (one overseeing basic research, the other applied), each with some 35 members. Six members of each commission sit on the committee itself, with the chairman and a handful of ministers from the government departments whose research spending has now been handed to KBN.

KBN's chairman, Witold Karczewski, a neurophysiologist from the Academy of Sciences Institute of Medical Research in Warsaw, is an energetic advocate of reform (he heads TPKN, the Polish Society for the Advancement of Sciences and Arts, which was banned under the martial law of the early 1980s as an 'anti-socialist' organization). His priority

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

*Hanging on — Olszewski became prime minister before Christmas but for how long?*

time that has elapsed since 1989 has been well used. In the last two years, the machinery of Polish government support for research has undergone a rapid transformation. With academics flocking into

COMMERCIALIZATION

## 'Happy accident' for technology transfer

*Warsaw*

KBN may have a monopoly on state funding for basic and applied research, but under Polish budget law, the committee cannot provide grants to encourage its commercialization. Given the dismal record of Polish industry in technology transfer, this is a cause for concern. But thanks to what Jan Krzysztof Frackowiak, under secretary of state at KBN, calls "a happy accident", there is another organization ready to step into the breach.

The Foundation for Polish Science (FNP), a new body aiming to stimulate technology-based business, owes its existence to some \$100 million left unspent from the Polish science budget of 1990. Despite the government's present shortage of funds, FNP has been allowed to invest the money. Part of the earnings

will be spent on grants for technology-related projects of social value — FNP has already established a scheme to provide equipment for projects in child health. But the foundation also intends to introduce loans to encourage technology transfer, for researchers or industrialists able to produce cogently-argued business plans. Frackowiak, who also serves as chairman of FNP's governing board, expects these loans to carry a 30–40 per cent rate of interest — the commercial rate is 50 per cent.

Frackowiak acknowledges that few Poles have experience in producing the business plans required by FNP, and that finding people with sufficient expertise of business and technology to assess the applications will also be difficult. But he remains optimistic: "We will try to teach one another." □

THIS brief survey of Polish science was compiled by Peter Aldhous, who would like to thank all the researchers and officials interviewed for their hospitality. Maciej Zylicz deserves particular mention, for arranging a full and interesting tour of Gdansk. Thanks also to Iain Macpherson, from the British Council office in Warsaw, and Jerzy Tarajkowski, from the Polish embassy in London, for their help in arranging my visit. □



## Grants scheme struggles ahead

*Warsaw & Gdansk*

WHEN Witold Karczewski was appointed to head KBN, in Spring 1991, his immediate challenge was to supervise a new western-style system of competitive research grants to individual researchers. After decades of funding by bureaucrats, who poured money into institutes seemingly oblivious of the quality of the work they were supporting, the new system has literally been built from nothing. Not surprisingly, given the flood of applications (around 10,000) in the first grant round, Karczewski and his colleagues on KBN admit to teething troubles.

The biggest problem has been finding reviewers to vet the applications. KBN took advantage of the election which defined its membership by asking voters to fill in a brief questionnaire about their own work, which provided the beginnings of a database of referees. But in many disciplines, the size of the Polish scientific community is small — a common complaint is that there are only two or three Polish researchers capable of reviewing a particular grant proposal. Karol Taylor, from the University of Gdansk, and a KBN member responsible for distributing grants in molecular biology, says that most of his outside reviewers were fairly uncritical, presumably because they knew the applicants. In the end, most of the difficult decisions that led to the final molecular biology ranking list were made by a committee of 10 biologists appointed by Taylor to help with the review process.

One solution — at least until Poland's scientists have learned to apply some harsh objectivity to peer review — will be to involve referees from abroad. The US National Science Foundation and the North Atlantic Treaty Organization

have offered their referee databases to KBN, but for the committee to employ foreign reviewers, Polish scientists must begin to submit their grant proposals in English. KBN is encouraging applicants to do so, and in fields where Polish science is relatively strong, such as physics, mathematics and neurobiology, many applicants are heeding this request. But Karczewski says it will take time before most Polish researchers, many of whom fear that their work would be viewed as "parochial" by foreign referees, follow suit.

The sheer scale of the first grant round gave KBN little opportunity to check the validity of many applications. Karczewski says he has exercised his veto over KBN's decisions to block the award of several grants, after reading applications that he was sure were attempts to defraud the committee, rather than genuine research proposals. But as the system settles into a routine of twice-yearly grant rounds, and the number of applications should become more manageable, this problem should ease.

Karczewski's biggest disappointment is that the cuts he was forced to make to the 1991 science budget hit the competitive grant system hardest of all: around \$50 million was awarded, a quarter of the figure that KBN had planned. (By the time that the budget cuts were announced, most of KBN's other spending had already been committed.) Nevertheless, Karczewski is determined that the competitive grant scheme should grow to become the cornerstone of Polish science funding. He would like grants to consume half of the national science budget, rather than the single-figure percentage they represented in 1991. □

is to ensure that KBN's scarce resources are given to the best institutes and researchers in the country. A new system of competitive individual research grants, which struggled to life last year (see above), will help, but Karczewski also intends to be more discriminating when handing out money to Poland's research institutes, which now receive their core budgets from KBN. Under the old system, government spending on research was dominated by large block grants to institutes and laboratories (with little consideration of the quality of research going on inside) rather than awards for specific projects. This was still largely the case for the 1991 science budget, the first handled by KBN, but Karczewski promises changes in 1992.

Karczewski's guide, as he tries to introduce some selectivity into Polish science, is a comprehensive assessment of

all Poland's research centres — the 80 institutes under the Academy of Sciences, the applied research centres linked to various ministries, and university departments. Karczewski believes that the ranking lists which have now been drawn up — based on the numbers of PhDs and full professors at each institute, and their publication records — are uncontroversial. "Very few people have protested", he says.

This year, Karczewski promises to concentrate funding in centres near the top of the ranking lists. He freely admits that some institutes will be forced to close. But for the time being at least, it seems that scientists employed on basic research have relatively little to fear: most of the institutes of the Academy of Sciences, the bastion of Polish basic research during the Communist era, were rated highly by KBN. And with the committee domi-

nated by researchers from the universities (70 per cent of Poland's researchers are based in institutes of higher education), KBN has an inbuilt sympathy towards basic science.

For industrial researchers, however, the prospect of unemployment is all too real. Jan Krzysztof Frackowiak, under-secretary of state at KBN, is scathing about the quality and relevance of the work in the 150 or so laboratories linked to the industry ministry. Less than 30 per cent of their work has ever been used by Polish industry, he estimates. Karczewski predicts that as many as 30 of these institutes may be shut down in 1992, after their funding from KBN dries up — making several thousand researchers and technicians unemployed, if they cannot find funding from elsewhere. Karczewski argues that industrial research should be paid for by its users, rather than by the state. But given the virtual collapse of Polish industry (most of which is, in any case, still state-owned) there is little hope of a new wave of research contracts to reinvigorate Poland's industrial laboratories.

Despite KBN's determination to end state subsidies for sub-standard industrial research, Karczewski would like to identify a number of promising industrial sectors where government support for research could pay dividends. Asked to name some examples, he suggests the production of small aircraft, and medical microelectronics. But at present, Karczewski can do little until the government produces a long term industrial policy. With its resources badly stretched, KBN cannot afford to start picking winners on the industrial scene, only to find that the government chooses a different range of enterprises as worthy of financial aid. It is here that the lack of a strong, stable government will be a problem, although a more serious threat to the reforms planned by KBN seems for the time being to have receded. One possibility, as Poland's centre-right political parties manoeuvred to build a government before Christmas, was that the chairmanship of KBN (which carries ministerial rank) would be given to a career politician from one of the minor parties, lacking Karczewski's experience in science policy. But Karczewski and Frackowiak's shrewd strategy of keeping on the sidelines, maintaining friendly relations with all of the major political parties, without seeming to have become too close to any one of them, has been successful, and Karczewski remains in office.

But the reform of Polish science may be less important to the new Polish parliament than for its predecessor. Most of the academics who entered the Senat and Sejm in 1989 have now returned to research: before the 1991 elections, fully



Chairman of KBN, the state Committee for Scientific Research, Witold Karczewski faces further budget cuts in 1992.

one quarter of the 100-seat Senat boasted the title of professor, but only half a dozen of these academics remain. Most of the rest simply decided not to run this time around. Entry into parliament may have seemed the logical next step for many Solidarity activists from academia back in 1989, but as partisan party politics became the order of the day, the realization that "I'm a scientist, not a politician" was common. For the new parliament, the progress of legislation designed to continue the science reforms initiated with the Act that established KBN is unlikely to be a priority. The main outstanding task is deciding a new legal framework for the Academy of Sciences, which has lost its role of distributing a large portion of the national science budget, but retains a large and mostly redundant bureaucracy, according to its critics (see page 388).

Meanwhile, the main concern for Polish researchers will be money. Under the Mazowiecki government, Polish science experienced a temporary period of prosperity — many university laboratories house expensive items of equipment (centrifuges, computers and the like) purchased in 1990. But most researchers now say that they have less to spend in cash terms than they did under the Communists. This statistic is slightly misleading, however, as the fact that the zloty is now a genuinely convertible currency (with a reasonably stable exchange rate of around 11,500 to the US dollar) has made a huge difference. Whereas western equipment and reagents could only be purchased before from the minuscule quantities of hard currency given to each laboratory (typically only a few thousand dollars for an active university research department), researchers

can now spend as much of their money as they like on supplies from the West.

Nevertheless, the financial hardship in Poland's laboratories is real. Karczewski was last year forced to cut the Polish science budget by more than 30 per cent, from the planned figure of around \$1,000 million. And with the national budget deficit still a huge problem, spending on research in 1992 will be squeezed still further. Ironically, the present shortage of money may actually make Karczewski's difficult job a little easier, although this will bring little comfort to the average Polish researcher. Even in the Academy institutes and in the universities, the few Polish researchers who have built solid international reputations say that a sizeable number of their colleagues (as many as half, according to harsher judges) are unproductive, and should be

INTERNATIONAL COLLABORATION

## A lifeline from abroad

*Warsaw*

**WHEN the states of Eastern Europe broke free from the influence of Moscow, Poland's scientists had something of a head start in tackling the urgent business of reform: even under the old Communist regime, most Polish researchers were able to travel and work in the West with relative ease. As a result, reform has been driven by a scientific community well aware of the limitations of the Soviet model of organizing research. (And in any case, this model was less rigidly imposed on the Poles than in many of Moscow's other former satellite states.) The fruits of this freedom of travel — the large number of ongoing collaborative projects linking Polish and western laboratories — may also serve to help cushion Polish science from the worst effects of the deep recession that is now gripping the country. Even if research groups do not have the money to buy vital equipment, they should still be able to do part of their work in their collaborators' laboratories abroad.**

The most common destination, for Polish scientists looking to work in the West, is the United States. Despite the closer proximity of Europe, many Poles find it easier to arrange work in American laboratories — where personal contacts are the key factor, and government bureaucracy is less of a hurdle. US funding agencies are also taking the lead in extending their reach to help support Polish science, presumably convinced that US researchers have much to gain from collaborating with the cream of Poland's researchers. The National Institutes of Health, for example, have recently started awarding 'shuttle grants' for Polish scientists, which pay for a series of two to three month visits to a laboratory in the

replaced. The current enforced financial stringency may allow KBN to prune some of the 'dead wood' from the Polish research community with a minimum of protest. But even some among KBN's supporters fear that the committee's new selective funding criteria may not ensure that it is Poland's least productive researchers that are the first to join the unemployment benefit queues. Andrzej Ziabicki, from the Academy Institute of Fundamental Technological Research in Warsaw, and a colleague of Karczewski in the leadership of the reformist TPKN, is worried that many of the least active researchers will have plenty of time on their hands to put together superficially appealing proposals for KBN funding, which if not reviewed thoroughly, may squeeze out proposals from their more deserving colleagues. □

United States. These are particularly popular with the new generation of Polish science policymakers, offering strengthened ties with the US biomedical research community while minimizing the likelihood that the grant holders will leave Poland for good.

But, with many European agencies also showing a renewed interest in Polish science, Britain remains a notable blackspot. Most Western countries have by now dropped visa requirements for visiting Polish nationals, but the British government still insists on subjecting Poles who wish to enter the United Kingdom to an obtrusive, often highly personal, inquisition, according to Polish scientists who have been through the process. Visas are often delayed, and occasionally denied. Understandably, few Polish researchers look first to Britain when trying to extend their network of international contacts.

As in the rest of Eastern Europe, a possible 'brain drain' of scientists to the West is a concern, although the fact that a temporary stay in a western laboratory has long been a feature in the career of many Polish researchers provides some reassurance that visits abroad do not necessarily lead to permanent emigration. The failure to attract the brightest Polish students into a career in research offering only minimal financial rewards is a far greater worry. Indeed, as a result, Polish laboratories may become the end point for another brain drain. Aleksander Zamojski, from the Academy of Sciences Institute of Organic Chemistry in Warsaw, and until this month president of the Polish Chemical Society, says that "we're thinking quite seriously of importing Russians and Chinese" to fill research posts spurned by young Poles. □



# Monopoly threatens variety

Warsaw

IRONICALLY, Poland's move away from a centrally planned economy has been accompanied by a greater centralization of decision making for science. With the entire Polish research budget in the hands of the state Committee for Scientific Research (KBN), some researchers are concerned that this monopoly will result in the neglect of important areas of science.

According to Jan Steffen, from the Maria Skłodowska-Curie Institute of Oncology in Warsaw, the policies of KBN threaten to undermine Polish clinical research. For fifteen years, Steffen's institute was responsible for coordinating Poland's cancer research: collecting epidemiological data, conducting basic research, organizing multi-centre clinical trials of new treatments, and awarding grants to researchers in medical academies throughout Poland. But the programme, totalling some \$10 million a year, has now been closed down.

Steffen is proud of the programme's achievements — for several types of cancer, survival rates in Poland are now approaching those in the United States and western Europe. But he fears that KBN's concentration on small grants to individual researchers, at the expense of large research programmes, will interrupt this progress. Already, says Steffen, the Skłodowska Curie institute's links with its collaborators in the medical academies are becoming "a little loose". Given the poor coordination between basic and clinical research groups at most of the medical academies, Steffen believes clinical oncology will suffer as a result.

Steffen understands KBN's reluctance to support large research programmes. The attempts by the Communist government to organize industrial research on a similar basis in the 1970s are widely regarded as a disaster. "But we have to find the proper balance," he says. At present, Steffen believes there is a tendency to reject all policies associated with the Communist system, even those that are "politically neutral".

Although Steffen voices his concerns louder than most, the wisdom of having a single committee, dominated by university professors, distributing the entire national science budget is questioned by a number of researchers. Among KBN's fiercest critics is Leszek Kuznicki, scientific secretary of the Academy of Sciences, who argues that the science spending would have been protected from the savage cuts it experienced in 1991, if the national science budget had remained spread over a number of ministries, rather than being rolled together into a single, easily targeted budget item. Kuznicki's point of view may be coloured by

the fact that the Academy's scientific secretary was formerly responsible for a large proportion of the funds now under KBN's control, but as the financial squeeze on Poland's researchers continues, he may be given a sympathetic hearing.

Nevertheless, most researchers seem reasonably happy with KBN's handling of basic science, although Steffen's concerns about clinical research are echoed in other applied disciplines, including agricultural and environmental research. Andrzej Ziabicki, from the Academy of Sciences Institute of Fundamental Technological Research, and a leading voice in the reformist Polish Society for the Advancement of Sciences and Arts (TPKN), believes there is a case for setting up a number of separate government agencies to coordinate research in these areas.

Witold Karczewski, chairman of KBN, acknowledges the concerns about the demise of the large state research programmes. But he says that KBN is offering packages of individual grants to unite groups working on related projects, which should allow for a degree of coordination. And Karczewski explains

SCIENCE ADMINISTRATION

## Where now for the Academy?

Warsaw

ALMOST everyone involved with Polish science agrees that the Polish Academy of Sciences has to change, but how this will be achieved is far from certain. The basic problem is simple: the structure of the Academy has changed little since it controlled much of Poland's budget for basic research, a function that has now been stripped away.

Most Polish researchers agree that an appropriate role for the Academy would be as a respected body of academics, available to provide impartial advice to the government — something along the lines of Britain's Royal Society, or the US National Academy of Sciences. But carving out this role will require the dissolution of much of the bureaucracy that still surrounds the 300-plus members of the Academy itself. Witold Karczewski, chairman of the state Committee for Scientific Research, which has taken over the Academy's responsibility for distributing research funds, notes enviously that the Academy seems still to have more administrative staff than his own committee. "They have very little to do," he says.

But even without the inherent conservatism of the academy — its members, though elected, average well over 60 years of age — there are serious obstacles to reform. The main complication is the

fact that there are good reasons to reject Steffen's suggestion that some of the programme grants should be reinstated, and that the relevant ministries (health, agriculture, and others) should have a supervisory role.

Before science funding was transferred to KBN, says Karczewski, the individual ministries applied little selectivity to ensure that the best groups received funding, and some are also thought to have diverted a proportion of their science budgets away from research. The problem with ceding some control over the national science budget back to the ministries, says Karczewski, is that many of the middle-ranking bureaucrats responsible for science under the Communist regime are still in their jobs. "It's not easy to change their mentality," he says.

Nevertheless, Karczewski is sympathetic to Ziabicki's idea of setting up a number of new research agencies alongside KBN. In time, he hopes to achieve this, provided that they can be populated with scientists, rather than bureaucrats. No clear plans have yet been drawn up, but Karczewski is looking closely at the British system — where the five research councils each support a different area of science, and are largely independent from their paymasters in government.

fact that the Academy was more than just an academic society and a channel for government research funding. Following the Soviet model, the Academy is also Poland's leading scientific publishing house, and ran 80 research institutes, most of them in the capital, which are generally agreed to house some of the strongest research laboratories in Poland. Discussion of the Academy's future inevitably becomes bogged down in disagreements over the fate of its institutes, still nominally controlled by the Academy's scientific secretary, Leszek Kuznicki.

It is generally agreed that Polish higher education would benefit from a greater involvement of Academy researchers in university teaching. Radicals such as Andrzej Ziabicki, from the Academy Institute of Fundamental Technological Research in Warsaw, would like to convert some Academy institutes into semi-autonomous research centres, closely associated with the universities, and helping out with teaching. But as this would probably involve the laying off of a substantial number of Academy institute and university staff, Ziabicki says there is little enthusiasm for his idea among the research community. And in any case, tinkering with Academy institutes will have to wait until the uneasy relationship between the institutes and their parent body has been resolved. □

Many researchers in the institutes say that the role of the academy's central office in the Palace of Science and Culture in Warsaw's city centre in governing the institutes has now diminished to the point that it is now time for the institutes to go their own way. The Academy's control over its institutes was formidable: until last year, Academy institute researchers wanting to travel abroad had to collect their passports from the Academy's offices. But those days are now gone, and institute directors now have little need to keep contact with Kuznicki, in managing their day-to-day affairs.

One idea that is much discussed is for the institutes to be reconstituted as inde-



*The Academy occupies part of the monumental Palace of Science and Culture.*

pendent national laboratories, perhaps modelled on the institutes of the French Centre National de la Recherche Scientifique. But Wladyslaw Findeisen, one of the Academy's four vice-presidents, and as head of the Senat Commission on Science and Culture, one of the legislators who could help bring this change about, says that conflicting claims over the Academy of Sciences' name will prove a problem. Largely due to the relative strength of the Academy institutes in research, the Academy's title is seen as a badge of scientific quality. But as a member of the Academy proper, "I don't want people in independent institutes to take our name," says Findeisen. He is working hard to push through a new law to clarify the legal status of Polish research centres, which will make some private institutes eligible for state support — but this law has nothing to say about the Academy's institutes.

With little enthusiasm in parliament for rapid legislative reform, it seems that the main hopes for change must lie with the Academy itself. Critics say that this is a recipe for inaction, pointing out that Academy members have little incentive to press for reform. If the Academy does

become simply the Polish equivalent of the Royal Society, there seems little justification for the substantial financial remuneration now given to its members. "They hate the Academy, but they love the money," is the harsh judgement from one of Poland's leading palaeobiologists, himself based in an Academy institute, speaking of the university professors whose salaries can be increased by 50 per cent or more by their membership of the Academy.

Kuznicki, who as scientific secretary is the real centre of power in the Academy (the post of president, held by the historian Aleksander Gieysztor, is largely a decoration), rejects the criticisms of the academy as a body resistant to change. Many of the Academy's critics are, he suggests, "frustrated professors" who have tried and failed to become members themselves. Since taking over as scientific secretary in January 1990, Kuznicki says that he has cut the Academy's adminis-

trative staff by more than 70 to 190 people. Those remaining, and the stipends paid to the Academy members themselves, he believes are justified by the Academy's continuing functions. Apart from the institutes, Kuznicki points out that the Academy is responsible for five laboratories abroad (including an Antarctic research station), six experimental farms, and numerous agreements for scientific cooperation with foreign scientific organizations. Gieysztor is less bullish, accepting that "substantial changes" are needed.

Whatever the future shape of the Academy, it now has some healthy competition as Poland's premier academic body. Already, several of the historic learned societies dissolved to make way for the Academy in the early 1950s, such as the Warsaw Society for Sciences and Arts and the Polish Academy of Knowledge in Krakow, have been restarted, and are flourishing. □

#### UNIVERSITY STRUCTURE

## Shoestring budget slows reform

### Gdansk & Warsaw

FOR Poland's universities as for the rest of society, the prospects for a better future are inextricably tied to the performance of the country's struggling economy. Asked to identify his priorities, Zbigniew Grzonka, rector of the University of Gdansk, puts invigorating the university's inefficient administration near the top of the list. Since taking over as rector in December 1990, Grzonka has found most of his time taken up with day-to-day administrative chores, rather than his plans for the university's future. The basic problem is money, and as Poland's economy falters, the universities cannot expect a rapid solution. The wages of university administrative and secretarial staff are pitifully low, and training in the past has been non-existent. When Maciej Zylicz joined Grzonka's team as prorector for scientific affairs, he says that many of his office staff would not even answer the telephone.

The salary problem also affects academics. A university professor earns around \$200 a month. Rents are low, but the prices of food and fuel are rising, and many professors must take second jobs. Not surprisingly, few students are opting for a life in teaching and research. "My brightest students tell me: 'Sorry, but I have to support my family'," says Tatiana Klonowicz, a University of Warsaw psychologist who supplements her income by interpreting at international conferences. This 'internal brain drain' is one of the main concerns of the education ministry, and of the state Committee for Scientific Research. But until the economy improves, scope for government action is limited.

Poland also has one of the lowest rates of

participation in higher education in Europe. In the absence of money to support a large increase to the country's 300,000 student population, the loss of academic staff and even the closure of some of Poland's institutes of higher education seems a distinct possibility, if the country's low ratio of teachers to students (around 1:6) is to be brought in line with the European norm, of greater than 1:10. Although Poland has more than 100 such institutes, many of these are small, specialized schools, bearing little resemblance to a Western university.

One development that could invigorate the Polish university system would be the increased mobility of academic staff, who now tend to spend their careers tied to a single institution. The problem is the lack of a free market in housing: rented accommodation becomes available only after years, and mortgages are an unknown phenomenon. Polish researchers now find it easier to work abroad than to move between laboratories in their own country.

Some progress should be less constrained by the economy. University rectors hope to modernize the Polish system of undergraduate teaching, which is based around five-year masters degrees. Shorter courses leading to a western-style bachelors qualification would be more appropriate in today's Poland, and could attract more students, the rectors believe. For the universities to reform their courses, laws that require a masters degree for entry into many professions must be changed. But given the heavy demands on the Polish parliament's time, the widespread introduction of bachelors degrees may have to wait for several years. □



# Democracy a mixed blessing

## Gdansk & Krakow

DEMOCRACY has run riot in Poland's universities. Under a law passed last year, the rectors who head Poland's top institutions of higher education must be elected by university staff and students, and will be called to account in repeat ballots every three years. Although faculty members did have some influence in the appointment of rectors previously, the final choice lay with the government. Not surprisingly, the first free elections in 1990 brought a host of new faces to the pinnacle of Polish academic life.

At the University of Gdansk, Zbigniew Grzonka confesses that he was initially a reluctant occupant of the rector's office — before being chosen by his colleagues to run the university, Grzonka was planning to travel abroad to continue his research in chemistry. Grzonka is now working energetically to purge the university of the legacy of Communism. But ironically, last year's legislative reforms, which guaranteed university autonomy, are proving to be an obstacle, as well as an opportunity.

Grzonka is trying to coax those departments where teaching and research was stongly influenced by the Communist ideology — such as the economics faculty — back towards the academic mainstream. But the reforms of 1990 have given a large degree of autonomy to faculty deans, who are also elected. Not surprisingly, in departments dominated by ex-Communists, Grzonka's calls for reform are going largely unheeded. "Now the freedoms are working against us," says Maciej Zyliec, prorector for scientific affairs. Coming from a Solidarity strike organizer whose own fight for freedom brought exile from Poland for two years in the early 1980s, Zyliec's sense of frustration is felt deeply.

Despite the problems, Grzonka and Zyliec are making progress. Grzonka controls the university's budget, so can reduce funding to departments that fail to reform their teaching. And he has also succeeded in setting up an independent western-style business school, where graduates will receive degrees from the University of Strathclyde, in Scotland. When Grzonka can be sure that the school will not come under the influence of the Communist old guard, he hopes to integrate it into the university.

Andrzej Pelczar, rector of the Jagiellonian University in Krakow, finds himself in a more fortunate position. Unlike the University of Gdansk, which was established by Poland's Communist government in 1970, the Jagiellonian has a strong tradition of independence which helped the university to continue to manage its internal affairs free from

excessive political interference through most of the post-war period. (Founded in 1364, the Jagiellonian is one of Europe's oldest universities.) "Our university was not broken at all. We saved our identity," says Pelczar.

Pelczar stresses that the democratization of Poland has not brought about revolutionary changes at the Jagiellonian — they simply were not needed. But the new political climate will allow Pelczar to tackle some undesirable hangovers from the Communist era. In the 1950s, Poland's universities were shorn of their medical schools, which were reconstituted as separate medical academies, linked to the health ministry. The Jagiellonian is expected to be the first Polish university to welcome its medical school back into the fold, perhaps within the next eighteen months. Both institutions should gain from the move. Medical graduates will benefit by receiving

## BIOTECHNOLOGY

# Entrepreneurs look for customers

## Gdansk

"I'm not the type of person who can wait for money to drop from the sky," says Anna Podhajska, of the University of Gdansk's department of microbiology, explaining her efforts to commercialize the work of her research group. In California, it may be common for university molecular biologists to double as biotechnology entrepreneurs, but in Poland, Podhajska is one of a rare breed.

Even before the Communist government's demise, Podhajska was beginning to interest western biotechnology companies in the restriction enzymes, plasmid DNA and other products produced by her group. Until now, her earnings have had to be in the form of reagents and other payments 'in kind', because the university could not accept cash payments on behalf of its employees. But all that is set to change: the university has set up a private foundation, numbering President Lech Walesa among its founders, which will for the first time allow Podhajska to earn cash. The French company Appligene, and Life Technologies of Maryland, are poised to become her first customers.

Podhajska's enthusiasm is infectious. Visitors to her university office are swiftly whisked out of the department, and into a neighbouring building housing her pride and joy: the new biotechnology laboratory that is the centre of her commercial operation. Finding space for Podhajska's new venture was difficult, but in 1989, university administrators agreed that she could convert an old kitchen, which for-

degrees from an internationally known university, rather than a relatively obscure academy. And the return of medicine to the Jagiellonian will bring it in line with top flight universities abroad. At present, Pelczar finds it difficult to explain to his counterparts abroad how a leading university does not have a medical school.

As the depressed state of the Polish economy brings financial hardship to the universities, the Jagiellonian may be better placed than most. Pelczar hopes that his university's international reputation will attract funds from abroad — already the Mellon Foundation, a US charity, has paid for computerization of the university library. But the Jagiellonian also has some earning potential, in the form of a new conference centre, complete with a 160-bed hotel. Although the centre has taken 20 years to build, during which time other development plans were severely restricted, it is now fully booked for the year ahead, and will be a important source of funds. □

merly served an austere adjoining student residential block (dubbed 'the Hilton' by its inmates). Luckily for Podhajska, her requests for money to buy equipment coincided with a temporary period of relative plenty for Polish scientists, under the initial Solidarity government of Tadeusz Mazowiecki, and the laboratory was completed in late 1990.

After returning some of her earnings to the university (part of the deal that won her the laboratory space was that a proportion of any profits would help with the Hilton's upkeep), Podhajska's priority is to reward her young research assistants, supplementing the \$100 a month they get from the university. "I'm here to make their lives decent," she says.

With the deep recession in the Polish economy squeezing spending on science, Podhajska's example may be a useful one for others to follow — and loans from the new Foundation for Polish Science (see page 385) may help other academics with commercial ambitions. But after over 40 years of Communist rule, few researchers share Podhajska's entrepreneurial spirit. And even for Podhajska, a commercial strategy is not without its costs. Until 1990, she kept her own laboratory bench space in Gdansk. But with an increasing amount of her time spent on administration, and in courting western biotechnology companies, Podhajska finds she has time to work on her own basic research only during her regular visits to the United States. □

# Russian science faces economic crisis

The Russian part of the ex-Soviet scientific enterprise is facing its economic crisis with fortitude. But these are only early days, and there is worse to come.

## Moscow

A now-fashionable sport among Russian researchers is the ironical calculation of personal salaries in US dollars. One laboratory director on Thursday last week announced that his salary, now 1,000 rubles a month, is "just \$10 a month". His purpose was to demonstrate that the present economic situation is self-evidently absurd. But he was out of date. The same week's auction of dollars appears to have set a rate of 210 rubles to the dollar. The director's salary is worth only \$5 a month. In one sense, of course, it does not matter. The prices of goods may have increased — meat prices have gone from 7 to 70 rubles a kilogram for example — but they are still denominated in rubles. Even so, with the economic reforms only a few weeks old, and with three months of winter (so far mercifully mild) to come, it is mystifying how people hope to plot a strategy for survival. There will be some relief among researchers that the three-month budget agreed between the Russian government and the Russian Academy of Sciences allows for all salaries to be increased by a factor of 1.9, but even that will leave many less senior people with salaries of only a few hundred rubles.

The situation creates endless minor humiliations. To buy a bottle of vodka, for example, three conditions must be satisfied: you must have a ration coupon, an empty bottle (to trade for the full one you hope to purchase) and there must be a vodka shop flush with a delivery of full bottles — not a predictable event. So people go to work carrying empty bottles in a plastic bag, just in case good fortune strikes the neighbourhood of their laboratory. Worse, because it is never possible to tell whether there will be vodka or only wine on sale, it makes sense to carry an empty wine bottle as well as one for vodka.

The economic reforms also affect the scientific enterprise more directly. Laboratories that used to enjoy the privilege of an allocation of hard currency have mostly now lost it. The result is that subscriptions to journals have almost everywhere been stopped. (The central library of the Russian Academy of Sciences is said to have abandoned its *Nature* subscription, for example.) For the same reason, overseas travel has become virtually impossible. It used to be that institute directors could buy air tickets from Aeroflot in rubles overvalued by the old artificial exchange rate. But a ticket to New York is now the

equivalent of ten annual salaries. Who will now ask to attend a conference overseas at the cost of ten colleagues' jobs?

Sadly, even the telephone will not be a convenient substitute, except for the laboratories with the foresight and clout to provide themselves with an e-mail link. Last week, it was almost casually announced that the cost of overseas telephone calls will be increased by a factor of 30, so that a call to a European destination will cost 180 rubles a minute and one to the United States half as much again. It is no wonder that people are musing ironically — as yet, there is no tinge of bitterness in the complaint — that freedom seems to have brought a degree of isolation from the international community even more complete than in the bad old days.

Surprisingly, shortages of laboratory supplies seem less pressing. Molecular biologists, for example, say that there are few difficulties in the supply of radiochemicals from Tomsk (the site of the first Soviet nuclear reactor), while common reagents such as restriction enzymes are also to be had with relative ease. Part of the explanation is that the Russian contribution to the Human Genome Project, one of 14 projects directed centrally by the government, has been supported generously over the past few years, and has chosen to invest heavily in infrastructure. But other more specialized reagents are a much more serious headache: people seem to have become at once skilled and shameless at borrowing scant supplies from colleagues and competitors abroad.

Just a few weeks into economic reform, there is necessarily a sense of unreality about. Often, people have a few dollars saved from the past (but foreign denominated accounts held with the old Soviet Central Bank have been frozen since the bank announced in December that it had run out of reserves; they may be unfrozen again in mid-February, but on terms not yet made public). Now, there is a prospect that people will soon be faced with the loss of jobs. While the Russian Academy of Sciences has, for the time being, resolved that it will not put people out onto the streets, that does not apply to its subsidiary agencies, the army of people working, for example, in its publishing house, Akademiye Nauk, which lost 14 million rubles last year.

Different institutes will deal differently with the money problem. At the academy's Institute of Molecular Biology, for

example, the plan is that those able to win research grants and contracts from outside will have their salaries substantially increased, perhaps doubled. At some institutes, it is claimed, outside income may well amount to 30 or 40 per cent of total costs. The outside sponsors of research are often other government departments or even industrial enterprises. There may be some nasty shocks ahead if the government's declared intention to balance spending and revenues in the first quarter of the year means that the sponsors are less willing or able to continue their contracts.

At other institutions, bigger upheavals are intended. At Dubna, the cooperative research centre organized around the development of nuclear energy, but now well on the way to becoming a general physical research laboratory, former members of the consortium (including many eastern European governments) have declared their unwillingness to continue, but that gloomy news has been to some extent offset by the discovery that the membership of the old German Democratic Republic (GDR) has now been subsumed in that of Germany as a whole. (The Bonn government undertook to meet the external obligations of the GDR at the time of German unification.) Does that spell a more glamorous future for Dubna than anybody has yet imagined?

Only time will tell. Meanwhile, the government seems to have set about an energetic examination of how the laboratory might be reorganized. The difficulties are huge. Not the least of them is that the town of Dubna, in Upper Volga lakeland country, is a single-purpose company town whose residents consist simply of those required to run the laboratory. Anything like radical change will be a social as well as an economic upheaval.

Meanwhile, the managers at Dubna have taken the most conspicuously enlightened step so far by deciding to recruit a new director for Dubna on the international labour market. A formal advertisement will appear among the classified advertisements in *Nature* next week. But those who may be interested should send an application and a curriculum vitae to Mr Ivan Bortnik in Moscow on the following numbers: fax (095) 230 2460 or e-mail [nsenche@ypr01.jinr.dubna.su](mailto:nsenche@ypr01.jinr.dubna.su). The new Russia is plainly determined to rejoin the international community. Will its economic problems allow it to do so successfully?

**John Maddox**



# Growing old before our eyes

Achim Weiss

A REPORT in this issue<sup>1</sup> describes the first direct detection of ageing in a star. In general, stars last far too long for astronomers to expect to see any evolution in the course of a working lifetime — unless they are lucky enough to see one destroy itself in a rare supernova explosion. But observations accumulated over the past 300 years, presented by de Groot and Lamers on page 422 of this issue, show that the visual brightness of the supergiant star P Cygni has been steadily changing as the star grows older.

Stars are long-lived objects with ages as old as our Universe, and therefore they have always been symbols of eternity because, to a human observer, they never seem to change. Only since Eddington, who suspected that nuclear reactions are the stars' primary source of energy, and since Atkinson and Houtermans, who showed how the fusion of hydrogen to make helium could take place, has it been known that stars have to evolve, and that the timescale for this ranges from millions to billions of years (see Table). Fossil records on Earth confirm that, at least during the past several hundred million years, our Sun has not changed its brightness by more than one per cent. The theory of stellar evolution, developed in the past 40 years, very accurately predicts the lifetimes of stars for different stellar masses and evolutionary stages; but a direct, observational verification is almost impossible.

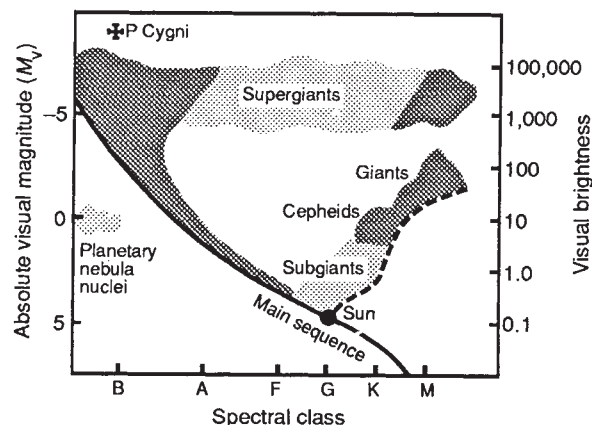
Up to now, astrophysicists have made use mainly of two indirect methods to measure the speed of stellar evolution. The first is statistical. If one observes a group of stars (preferably a globular cluster), the number of stars in different regions of the colour-magnitude (bright-

ness) diagram, which uses the two most directly measureable quantities, will reflect the times that stars spend in different stages of evolution (see figure). In particular, the number of red giants relative to main sequence stars will give the ratio of the lifetime of stars with hydrogen burning in a shell to the life-

time of those with core hydrogen burning. However, additional effects, such as the initial number of stars in every mass interval, also have to be taken into account. Another example of this approach is the so-called red-to-blue ratios in the colour-magnitude diagrams of our Galaxy and the Magellanic Clouds. Again, the relative numbers of cool (red) supergiants to hot (blue) ones serve as a test for the times spent in different phases. The situation is even more difficult here, as a wide range of stellar masses are involved, and the physics, most notably that of convection and mass loss, is accessible only through parameterized approaches<sup>2</sup>. Actually, the observations serve to tune these parameters.

Statistical methods, therefore, are good for confirming the magnitude of stellar lifetimes only, but involve additional uncertainties. Another approach is to use secondary properties of stars that change rapidly with the slow evolutionary changes. In particular, the periods of pulsating stars depend on their temperature and luminosity. The best known pulsating stars, the Cepheids, show relative period changes of the order of  $10^{-4}$  per century, which are rather easy to observe, as the periods themselves range from 4 and 40 days. Hofmeister<sup>3</sup> identified Cepheids as being evolved stars in a double-shell burning phase, which experience a loop in the colour-magnitude diagram that lasts for a few thousands years. From the computed relation between the period and the effective temperature and luminosity (the colour and brightness), she predicted that the period change should be observable within a few decades. Indeed, this has been the case<sup>4</sup>. Although additional uncertainties are involved, such as period changes due to other causes and the fact that those loops are rather sensitive to details of stellar structure, the observations essentially confirm the theoretical speed of evolution for such stars.

A similar approach is used for other classes of variables (R CrB, RR Lyr), but with less accuracy. A very peculiar case is FG Sagitta, the central star of a planetary nebula. This star is currently experiencing a thermal shell flash, which in turn affects the observed pulse period. So although astronomers believe in the correctness of the theoretically computed lifetimes of stars, an observation of the change in the colour and luminosity of a star is the direct measurement that has been lacking until now.



A schematic Hertzsprung-Russell, or colour-magnitude, diagram for the main classes of stars. For most of their lives, stars remain nearly unchanging on the main sequence, at the bottom right for cool, low-mass stars, at the top left for hot, massive stars. The giant and Cepheid classes are subsequent evolutionary stages. P Cygni has already evolved off the main sequence and is currently tracking horizontally to the right. (After Fig. 6.1 in ref. 5.)

de Groot and Lamers use observations made since 1700 of the star P Cyg. This is a massive star of several tens of solar masses with a surface temperature around 20,000 K. It is of fifth magnitude and can therefore be seen with the naked eye. Its absolute luminosity is more than  $10^5$  times that of the Sun. Its position in a colour-magnitude diagram (see figure) indicates that the star is in a phase of rapid transition, which lasts only several hundred years, between the main-sequence and the red-supergiant region. Unfortunately, the star is also an irregular, variable star showing outbursts and spectral changes which are due to dynamic events in the star's atmosphere that last years.

No matter which technical methods they chose, all observers could measure the brightness of P Cyg in the visual band. de Groot and Lamers proceeded in the following, strikingly simple way: the observations of the eighteenth and nineteenth centuries were visual estimates, with an uncertainty assumed to be about 0.3 magnitudes. The irregular variations account for an additional error of 0.2 magnitudes. Several observations were averaged to reduce these uncertainties. After 1898, astronomers used photoelectric photometry, which has the advantage of higher precision, but the disadvantage of a multitude of filter systems. All these measurements needed

TYPICAL LIFETIMES OF STARS

| Mass<br>(solar masses) | Phase                     | Lifetime<br>(years) |
|------------------------|---------------------------|---------------------|
| 1                      | Main sequence             | $10^{10}$           |
| 1                      | Red giant                 | $10^9$              |
| 20 (SN1987A)           | Main sequence             | $10^7$              |
| 20                     | Central helium<br>burning | $10^6$              |
| 3                      | Thermal pulse             | 2,000               |
| 9                      | Cepheid                   | 5,000               |
| 30 (P Cyg)             | Post main sequence        | 1,000               |

ness) diagram, which uses the two most directly measureable quantities, will reflect the times that stars spend in different stages of evolution (see figure). In particular, the number of red giants relative to main sequence stars will give the ratio of the lifetime of stars with hydrogen burning in a shell to the life-

to be transformed to the standard V magnitude. This way de Groot and Lamers constructed their Fig. 1 (page 422), which shows a convincing trend — even including errors — of a brightening of P Cyg at a steady rate of  $-0.15 \pm 0.02$  magnitudes per century, a sure sign of gradual evolution.

The safe assumption is that in its present state of evolution, the star's luminosity or bolometric brightness is almost constant, as theory predicts. In which case, the observed brightening can be due only to a change in its effective temperature (colour), which the authors compute to be 6 per cent per century. This can also be translated into a characteristic time for the stellar radius to grow by a factor of roughly 300 years. Indeed, stellar evolution theory predicts that the current evolution is due to an expansion of the envelope on a thermal (thus extremely short) timescale, expected to be about 500 years — close to that observed.

A more detailed comparison with computational tracks<sup>2</sup> also yields a difference of a factor of two (theoretical models evolve more slowly). However, GENE EXPRESSION

given the uncertainties in the physics used for the calculations and in the mass of P Cyg, de Groot and Lamers argue that this difference lies within the errors. The most important point, however, is that the timescale is indeed of the order of 100, and not 100,000, years or more, as is typical for other phases in the evolution of massive stars.

The ability to use the apparent visual brightness of a star to calculate the speed at which it is evolving may not seem to be a revolutionary finding, but its value lies in the simplicity of the method. Of course, direct observations of main sequence stellar evolution will be a bigger challenge. □

Achim Weiss is at the Max Planck Institut für Astrophysik, Karl Schwarzschild Strasse 1, W-8046 Garching, Germany.

1. de Groot, M. J. H. & Lamers, H. J. G. L. M. *Nature* **355**, 422–423 (1992).
2. Maeder, A. & Meynet, G. *Astr. Astrophys.* **182**, 243–263 (1987).
3. Hofmeister, E. Z. *Astrophys.* **65**, 194–225 (1967).
4. Fernie, J. D. in *Observational Tests Of The Stellar Evolution Theory* (eds Maeder, A. & Renzini, A.) 441–444 (Reidel, Dordrecht, 1984).
5. Aller, L. A. *Atoms, Stars and Nebulae* 3rd edn (Cambridge University Press, 1991).

## Dialogue with the cell cycle

Brenda J. Andrews

ON page 449 of this issue<sup>1</sup>, N. F. Lowndes and colleagues describe some remarkable results which bear on the question of how gene expression and DNA synthesis are coordinated with the cell cycle. Not only does their report demonstrate that the Start control protein encoded by the *cdc10*<sup>+</sup> gene of fission yeast is very likely to be a transcription factor, it also points to a conserved mechanism by which regulatory complexes in the upstream sequences of genes sense cell-cycle position.

Although the expression of most genes remains constant during cell growth, a small number show peaks of transcription at defined points in the cell cycle<sup>2</sup>. In particular, a recurring pattern is the restriction of expression of genes required for DNA synthesis to the late G1/S period. So far, the most impressive collection of genes exhibiting this pattern of cell cycle-sensitive transcription has been found in the budding yeast *Saccharomyces cerevisiae*; the expression of at least 17 genes required for DNA synthesis and metabolism seems to be coordinately induced during the cell-cycle, with maximal levels occurring precisely when a ready supply of the encoded enzymes is needed (see Table)<sup>1,3,4</sup>. Such a feat of coregulation would most logically be mediated by a common transcription factor providing a

link between the regulators of cell-cycle passage and the upstream regions of the various genes.

Two lines of investigation have supported this view. As sequence information about the coregulated S-phase genes became available, it was found that their upstream regions contain at least one copy of the sequence 5'-ACGCGT-3'. Because this sequence represents a recognition site for the *MluI* restriction enzyme, it was dubbed the *MluI* motif or MCB (*MluI* cell-cycle box)<sup>5</sup>. Studies of two genes, *TMPI* (*CDC21*)<sup>6</sup> and *POL1*<sup>7</sup>, which contain MCB sequences in their upstream regions, then established that the MCB element was capable of activating transcription in a cell cycle-dependent manner. In fact, the MCB sequence, in isolation, was found to direct cell cycle-regulated transcription of a reporter gene<sup>4,6</sup>, although the pattern of expression differed somewhat from that directed by MCB in its native context.

Meanwhile, yeast cells were being broken open in a biochemical effort to identify a factor capable of binding to the MCB sequence. A DNA-binding protein that has the correct recognition specificity, called MCB factor, has been highly purified, but as yet there is no evidence to link it with the cell cycle<sup>5</sup>. Another factor, DSC1 (DNA synthesis

control), was identified in crude cell extracts from its ability to recognize the MCB, in isolation or in the upstream sequences of the *CDC9* gene<sup>4</sup>. DSC1 was found to bind to synthetic MCBs with a cell-cycle periodicity that mimicked the activation conferred on a heterologous gene by isolated MCB sequences. Despite the differing periodicities conferred by isolated MCBs and those in context, these studies nonetheless established the MCB and DSC1 as excellent candidates for key components of a cell-cycle regulatory pathway; the circuit begins with an early event in the cell cycle and culminates in the coordinate induction of a group of genes necessary for DNA replication.

What of the protein or proteins of which DSC1 is composed? In a familiar performance on the stage of cell cycle research, *Schizosaccharomyces pombe*, the fission yeast, now arrives to share the spotlight with *S. cerevisiae*. The *S. pombe* gene *cdc22*, which encodes a subunit of ribonucleotide reductase, is the one DNA-synthesis gene known to be periodically expressed in fission yeast<sup>8</sup>. Johnston, Fantes and coworkers<sup>1</sup> noticed that the upstream regulatory sequences of *cdc22* contain two MCBs as well as several closely related sequences. They went on to show that isolated MCBs could function in *S. pombe*, as in budding yeast, to link transcription of a reporter gene to the cell cycle, and that a factor with binding properties indistinguishable from DSC1 was present in crude extracts of fission yeast.

Next, the group reasoned that a genetic handle on the identity of the *S. pombe* DSC1-like factor might be secured by analysis of the effect of known cell-cycle mutants on DSC1-like activity; if the factor were indeed necessary for the expression of genes essential for DNA synthesis, mutants in genes affecting the factor might be expected to be cell-cycle mutants that arrest at Start. Of the two cell-division cycle genes of fission yeast known to act in late G1 at Start, *cdc2*<sup>+</sup> is familiar to the readers of these pages as encoding the p34<sup>cdc2</sup> kinase whose activity is required to induce cell-cycle transitions at both Start and mitosis. The second gene, *cdc10*<sup>+</sup>, has been something of an enigma. But, with a series of incisive biochemical experiments, Lowndes *et al.*<sup>1</sup> have now established that its protein product is a component of the DSC1-like complex that they discovered in fission yeast.

An immediate consequence of this discovery is to establish that *cdc10*, whose activity is essential for passage through Start, is probably a transcription factor. As mentioned above, the only gene so far known to be regulated by the MCB-*cdc10* (DSC1) pathway is *cdc22*<sup>+</sup>. Unlike budding yeast, most DNA-



synthesis genes in *S. pombe* are constitutively expressed throughout the cell cycle and do not have MCB sequences in their upstream regions. Although *cdc10* mutants that are blocked at the restrictive temperature have much reduced levels of *cdc22<sup>+</sup>* transcripts<sup>1</sup>, it is not yet clear whether this alone can account for the inviability of *cdc10* mutants. Given the extensive use of the MCB-DSC1 system in budding yeast to link transcription to the cell cycle, it seems likely that *cdc10<sup>+</sup>* is regulating other genes in addition to *cdc22<sup>+</sup>* in fission yeast.

A second consequence of the discovery of the MCB-DSC1 pathway in *S. pombe* is to demonstrate the functional conservation of this regulatory system (fission yeast is related only distantly to *S. cerevisiae*). In particular, it raises the question of whether *cdc10*-like proteins are components of the DSC1 factor of *S. cerevisiae*, which is especially intriguing because two proteins in *S. cerevisiae* (SW14 and SW16) bear both functional and structural similarity to the *cdc10* gene product.

Like *cdc10*, SW14 and SW16 are both components of a transcription factor that links expression of target genes to the beginning of the cell cycle. Genes regulated by SW14 and SW16 include those encoding known and putative G1 cyclins whose activity is thought to activate the budding yeast version of the *p34<sup>cdc2</sup>* kinase at Start<sup>9,10</sup>. The resemblance in sequence between SW14, SW16 and *cdc10* is most significant (40–45 per cent similarity) within two 33-amino acid repeats first noted by Breeden and Nasmyth in their molecular analysis of the SW16 gene<sup>11</sup>. The so-called *cdc10*-SW16 or ankyrin motif has subsequently been found in a broad range of regulatory proteins<sup>3</sup>, and has been implicated in protein-protein interactions. In the case

of the yeast proteins, the motif may play an important role in their common mandate: to render transcription dependent on Start. The exact mechanism whereby SW14, SW16 and *cdc10* sense cell-cycle position is still a matter of conjecture; it has been suggested<sup>1,3,9,10</sup>, but remains to be shown, that these regulators may be directly phosphorylated by the *p34<sup>cdc2</sup>* kinase in G1 and that this modification leads to transcriptional activation.

So, are SW14 and/or SW16 in fact involved in coordinating expression of S-phase genes in budding yeast with the cell cycle? It is likely to turn out that they are, but several observations indicate that complications can be expected.

First, overexpression of *cdc10<sup>+</sup>* in *S. cerevisiae* cannot replace the need for SW16<sup>11</sup> (L. Johnston, personal communication). Second, the known target sequence for SW14 and SW16, which has been tagged the CCB (cell-cycle box), differs from that of *cdc10*-DSC1 (CACGAAA against ACGCGT). Because both the MCB and the CCB are clearly 'cell cycle boxes', the less ambiguous designation of SCB (SW14/16-dependent cell-cycle box) has been proposed for the CCB sequence (K. Nasmyth, personal communication). If SW14 and SW16 regulate transcription of S-phase genes in precisely the same way as their established targets, it would seem superfluous to have the MCB in one set of genes and SCB sequences in the other. Although both SW14 and SW16 are essential for SCB sequence activity, SW14 may be the component that recognizes the SCB and activates transcription while SW16 plays a regulatory role in the complex<sup>12</sup> (B. Andrews, unpublished). Perhaps the MCB and SCB pathways in *S. cerevisiae* are both linked to the beginning of the cell cycle by a common regulator (such as SW16) while unique components recognize the target sequence and confer pathway-specific regulatory properties on the responsive genes (for example SW14 for the SCB pathway, as yet unknown for the MCB pathway). As would be expected if this view were correct, the behaviour of the DSC1 factor in gel shift assays suggests that more than one component is present<sup>1,5</sup>.

Finally, there is the matter of whether homologues of *cdc10*, SW14 and SW16 regulate cell cycle-sensitive transcription in other organisms. Given the extensive conservation of regulators of cell-cycle progression, and the existence of analo-

## PERIODICALLY EXPRESSED S-PHASE GENES IN YEAST

| Gene                     | Gene product or function                     |
|--------------------------|----------------------------------------------|
| <i>S. cerevisiae</i>     |                                              |
| <i>CDC21</i>             | Thymidylate synthase                         |
| <i>CDC9</i>              | DNA ligase                                   |
| <i>CDC8</i>              | Thymidylate kinase                           |
| <i>CDC6</i>              | DNA synthesis                                |
| <i>POL1</i>              | DNA polymerase I                             |
| <i>POL2</i>              | DNA polymerase II                            |
| <i>POL3(CDC2)</i>        | DNA polymerase III                           |
| <i>POL30</i>             | Proliferating cell nuclear antigen (PCNA)    |
| <i>PRI1</i>              | Large subunit, DNA primase                   |
| <i>PRI2</i>              | Small subunit, DNA primase                   |
| <i>RNR1</i>              | Regulatory subunit, ribonucleotide reductase |
| <i>DPB2</i>              | DNA polymerase II subunit B                  |
| <i>DPC2</i>              | DNA polymerase II subunit C                  |
| <i>RAD54*</i>            | DNA repair                                   |
| <i>RFA1,2,3</i>          | Replication factor A                         |
| <i>S. pombe</i>          |                                              |
| <i>cdc22<sup>+</sup></i> | Ribonucleotide reductase subunit             |

\* G. Bezil, personal communication.

gous systems in distantly related yeasts, the answer is likely to be 'yes'. Some tantalizing similarities between known mammalian regulatory systems and the yeast circuits encourage this view. In particular, the site of action of the E2F transcription factor is similar to both the SCB and the MCB sequence (ACGCGAAA; S. Hiebert, personal communication). One potential target of E2F, containing E2F-binding sites of functional importance<sup>13</sup>, is dihydrofolate reductase, a gene required for S phase whose transcription is induced in late G1 (ref. 2). The scenario becomes all the more intriguing considering the involvement of E2F with regulatory events in G1. Further dissection of the regulatory machinery connecting transcription to the cell cycle promises to provide new insights into the mechanism of action of mitotic regulators in both yeasts and other organisms. □

Brenda J. Andrews is in the Department of Molecular and Medical Genetics, University of Toronto, Toronto M5S 1A8, Canada.

## Errata

■ We may inadvertently have misled readers of the item on invertebrate phototransduction (*Nature* **354**, 356; 1991) by not making it clear that patch-clamp analysis of isolated *Drosophila* ommatidia was first described by R. C. Hardie and colleagues (*Neuron* **6**, 477–486; 1991) in a study of voltage-activated potassium channels in adult photoreceptors. Hardie also reported (*Proc. R. Soc. B* **245**, 203–210; 1991) that calcium permeating the light-sensitive channels mediated deactivation of the light response.

■ In the account of the meeting on the biology of molecular chaperones (*Nature* **354**, 434; 1991), in one passage the word 'reported' was wrongly transmuted to 'found' during editing. The passage concerned should have read 'A. L. Horwich ... reported that an abundant protein of the thermophilic bacterium *Sulfolobus shibatae*, known as TF55 (for thermophilic factor 55), is very similar in sequence to TCP-1'. The principal investigator in the work concerned was J. D. Trent.

■ The News and Views on H<sub>2</sub> in extraterrestrial environments (*Nature* **353**, 502; 1991) omitted to mention its *in situ* detection in Jupiter's magnetosphere by D. C. Hamilton *et al.* (*Geophys. Res. Lett.* **7**, 813–816; 1980) in data obtained by the low-energy charged-particle instrument on the Voyager 2 spacecraft.

- Lowndes, N. F., McInerney, C. J., Johnson, A. L., Fantes, P. A. & Johnston, L. H. *Nature* **355**, 449–453 (1992).
- McKinney, J. & Heintz, N. *Trends biochem. Sci.* **16**, 430–435 (1991).
- Andrews, B. & Herskowitz, J. *J. biol. Chem.* **265**, 14057–14060 (1990).
- Lowndes, N. F., Johnson, A. L. & Johnston, L. H. *Nature* **350**, 247–250 (1991).
- Verma, R., Patapoutian, A., Gordon, C. & Campbell, J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7155–7159 (1991).
- McIntosh, E., Atkinson, T., Storms, R. & Smith, M. *Molec. cell. Biol.* **11**, 329–337 (1991).
- Gordon, C. & Campbell, J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6058–6062 (1991).
- Gordon, C. & Fantes, P. *EMBO J.* **5**, 2981–2985 (1986).
- Ogas, J., Andrews, B. & Herskowitz, I. *Cell* **66**, 1015–1026 (1991).
- Nasmyth, K. & Dirick, L. *Cell* **66**, 995–1013 (1991).
- Breeden, L. & Nasmyth, K. *Nature* **329**, 651–654 (1987).
- Breeden, L. & Mikesell, G. *Genes Dev.* **5**, 1183–1190 (1991).
- Blake, M. C. & Azizkhan, J. *Molec. cell. Biol.* **9**, 4994–5002 (1989).

# Snakes and female sexuality

G. A. Parker

LIBIDINOUS females? On page 440 of this issue<sup>1</sup>, Madsen *et al.* present new evidence why we may expect them to be so. The evidence and its interpretation is intriguing, not least because it runs counter to the longstanding view that, whereas males should be eagerly promiscuous, females, once inseminated, should be largely uninterested in further matings.

This traditional view of female sexuality is not simply male chauvinism and has a rational basis, as first defined by Darwin<sup>2</sup>. The costs of each offspring to a female are considerable, but the male normally contributes just a sperm. A new litter may take weeks to produce; an ejaculate may be replenished in hours. Such disparities result in males competing intensely over females<sup>2,3</sup>. For example, reproductive success in male *Drosophila* increases linearly with number of matings with different females, but female reproductive success is unaffected by number of matings after the first<sup>4</sup> — typically an ejaculate contains more than enough sperm to ensure fertilization, so no obvious gain accrues to a female by further matings. Clearly, the same autosomal genes will occur in both sexes, but we can talk independently of male or female interests (measured by male or female reproductive success) in the sense of the fate of a gene whose action is limited to one sex.

Although male interests are obvious, whether inseminated females should accept or reject further matings is controversial. Mating costs (time and risk) may well exceed benefits, but whether a female will show rejection depends on the outcome of an evolutionary battle of the sexes<sup>5</sup>. Even if forceful copulation can be prevented, courtship can be a war of attrition in which it pays to yield rather than to waste time rejecting. Successful rejection implies that females have won; acceptance implies either that they have lost or that multiple mating is advantageous for them as well as for males.

There are grounds, in principle at least, for thinking that re-mating might on occasion be beneficial for females. Possible 'environmental benefits' include the provision of food, protection and offspring care by males. Also proposed are 'genetic benefits', which relate to genetic variation among males. Walker<sup>6</sup> argued that female insects may use multiple mating for mate choice. For example, on meeting a male with 'better' genes the stored sperm can be swapped for his; or, if what will be 'better' genes cannot be predicted, multiple mating

might act as a form of genetic bet-hedging. But although the action of environmental benefits is uncontroversial and can be supported by direct evidence, that of 'genetic benefits' raises considerable debate, not least in relation to the maintenance of genetic variability.

In their fascinating field study of an adder population (*Vipera berus*) around Smygehuk, southern Sweden, Madsen

tion of stillborn offspring is reduced by multiple mating. (The population studied by Madsen *et al.* showed a surprisingly high proportion of stillbirths (31.6 per cent), which may be due to inbreeding depression.) If multiple mating is indeed a female strategy to facilitate the triumph of sperm from genetically superior males under sperm competition, the finding is exciting.

But there are some doubts. If rival ejaculates are similar in all but genes, the hypothesis of Madsen *et al.* demands a correlation between sperm success and offspring viability. It is not immediately

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Emotional entanglements — male and female adders courting.

*et al.* claim to have found empirical evidence that multiple mating increases female reproductive success through genetic benefits: genetically superior males appear to win more fertilizations under sperm competition<sup>7</sup>, raising the mean viability of offspring. There are no obvious environmental benefits. Males provide no parental care or nuptial gifts, and females can store sperm for months before ovulation.

Female adders mated between one and eight times (mean 3.7), and Madsen *et al.* say they avoid mating repeatedly with the same male thereby increasing the genetic diversity of the stored sperm. This claim is based on the (conservative) estimate that females mate three times, but assumes that the population of males around a female equals the average number which court her in a season (seven). With high mobility of one or both sexes, and an equal sex ratio in their population (mean 30), the mean number of males available could approach 15. The expectation of mating more than once with the same (randomly drawn) male is then 0.19 (less with a male-biased sex ratio), that is, close to the 0.16 they observed.

However, whether female adders mate randomly or avoid repeat matings is not crucial to the key result that the propor-

obvious why sperm carrying genes that increase offspring viability should out-compete rival sperm, unless the same deleterious mutations affect almost all aspects of the phenotype, from gamete to adult. (The geneticists' view is that a gamete's traits are determined by its diploid parent rather than by its own haploid chromosomes.) If this mechanism applies, we need proof that stillborn offspring tend to come from males whose sperm fare badly under sperm competition. Matings can be observed directly in the field, and paternity could be determined by genetic fingerprinting.

Another question is why females don't simply filter out 'poor quality' males by refusing to mate with them. This strategy would avoid the costs of multiple mating but, perhaps more important, it would eliminate all chance of zygotes from inferior males. Maybe females cannot discriminate between males, and rely on sperm competition as their only option. As regards the stillbirths, if inbreeding depression is the cause, the optimal mate choice will depend on the genotype of both parents. But to expect selection to produce mate choices which take account of the combined result of a female's genotype and that of her suitor is indeed to have faith.

What of alternative interpretations?

Ardea



No obvious environmental benefits accrue from multiple mating (no increase in offspring mass, litter size or total litter mass; nor do females gain higher fertility or lose less body weight during gestation). It could be that 'fit' females simply attract more matings, or that males exert a preference for 'fit' females. Or could the result arise from some ecological effect: perhaps population density (and hence the chance of multiple mating) is higher in 'better' parts of the habitat, with offspring of females in 'poorer' areas suffering adverse effects? The lack of correlation between multiple mating and such features as litter size and mass counts against such ideas, but does not invalidate them — for example 'poor' areas may contain some toxin which results in stillbirths. A further interpretation is that the effect arises through sperm ageing. Multiply mated females may have on average fresher (less defective) sperm than singly mated females.

A final possibility — fitting the authors' hypothesis and even providing a plausible mechanism for it — is that 'fit' males transfer more (rather than more vigorous) sperm. Multiple mating could thus be advantageous to females without any correlation between a sperm's vigour and an offspring's viability. A possible reason for 'fit' males ejaculating more sperm relates to male strategy under sperm competition: sperm expenditure is an evolutionary game in which the best strategy depends on the strategies of other males in the population<sup>8</sup>. Males may even increase sperm numbers if sperm competition is likely<sup>9,10</sup>. The cost of sperm to a 'fit' male may be less than that of an 'unfit' male, and he may therefore ejaculate more sperm<sup>8</sup>. In other words, 'fit' males can buy more tickets and win the raffle more often.

Whatever the exact mechanism underlying their result, if Madsen *et al.* are right that sperm competition allows a female to gain 'better' genes for her offspring we shall have a new reason why some females accept several matings, and a new role for sperm competition. I rather hope it turns out that they are correct. But for the time being judgement must be suspended. □

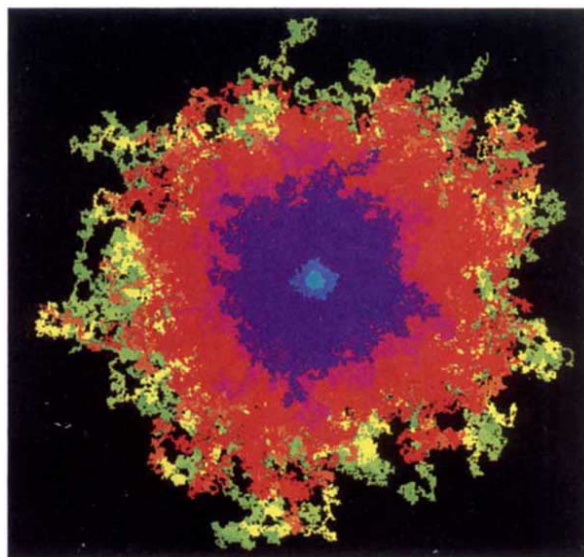
G. A. Parker is in the Department of Environmental and Evolutionary Biology, University of Liverpool, Liverpool L69 3BX, UK.

## New paths for random walkers

Michael F. Shlesinger

A LARGE class of problems involving noise, fluctuations, transport, relaxation and reactions can be cast in the form of a random walk. One area of active investigation<sup>1</sup> in random walks involves calculating the first passage time for a walker to achieve a goal, such as inducing a reaction by reaching a specific site or hitting another walker, and so on. Another active area involves calculating the behaviour of many random walkers. On page 423 of this issue<sup>2</sup> Larralde *et al.* combine both of these themes. They

large  $N$  and small  $t$ ) fill the volume, of radius  $t$ , explored around their origin. The number of distinct sites visited initially grows as  $t^d$ , where  $d$  is the dimension. This growth in the number of new sites visited slows down at intermediate times to  $t^{d/2}$  times a dimension-dependent function which has a weaker dependence on  $t$ . It is in this intermediate time regime where the geometry of sites visited becomes complex and interesting that the most important applications probably lie.



Territorial army: successive colours show the progress of 1,000 random walkers starting at the origin (centre) as they encroach on new territory over several successive periods. Note how the frontier of the explored territory is initially smooth, but later becomes rough. (Courtesy of Eugene Stanley.)

calculate the territory covered by  $N$  diffusing random walkers which start nearby each other. New territory is added when a site is visited for the first time. The complication arises that only the first walker to visit a new site counts, so to prevent overcounting one must keep track of the past history of all the walkers.

The authors' analytical and numerical work for the geometry of the new territory finds three distinct timescales of behaviour, including a fascinating transition in the roughening of the geometry of the sites visited (see figure). Only in the last regime when the walkers have moved far enough apart from each other would the results reduce to those of single-particle statistics scaled by a factor of  $N$ . At long times each walker visits new sites proportional to the number of steps taken (in three or higher dimensions). At early times many new sites are visited. After  $t$  steps the  $N$  walkers (for

One might have suspected that Jacob Bernoulli discovered most of the key ideas of random walks way back in the early 1700s. For example, he found that if two gamblers have  $R$  coins between them, and one coin is won or lost in each play of a fair game, then the mean number of plays  $\langle N \rangle$  before one player runs out of money is given asymptotically by  $\langle N \rangle = (1/D) R^2$ , where  $D$  is a constant. This was the 'first' first-passage-time calculation. It took Einstein, 200 years later, to move the brackets from time to space, to invent brownian motion with the mean square displacement of  $\langle R^2(N) \rangle = DN$ . For the above Bernoulli process, letting  $R$  and  $N$  be continuous, De Moivre, in 1756, in the 3rd edition

of his classic book *The Doctrine of Chances*, derived the first-limit theorem (for what we now call the gaussian probability density) for a player's fortune. So early on much was known about random-walk probabilities, and in principle everything about a random walk can be derived from knowing  $P_n(x)$ , the probability for being at a site  $x$ , after  $n$  steps.

The complexity of the many-walker problem studied by Larralde *et al.* is perhaps surprising in that the statistics of each walker are gaussian, and the walkers do not interact (influence one another). The complexity arises through the question which is asked. The authors' question involves a first passage time and a conditional probability for checking that when a walker reaches a new site, it is the first of the walkers to achieve this. It is the infinity of different questions that one can ask, even about simple random processes, that makes

1. Madsen, T., Shine, R., Loman, J. & Håkansson, T. *Nature* **355**, 440–441 (1992).
2. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (John Murray, London, 1871).
3. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 139 (Aldine, Chicago, 1972).
4. Bateman, A. J. *Heredity* **2**, 349–368 (1948).
5. Parker, G. A. *Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 123–166 (Academic, New York, 1979).
6. Walker, W. F. *Am. Nat.* **115**, 780–799 (1980).
7. Parker, G. A. *Biol. Rev.* **45**, 535–567 (1970).
8. Parker, G. A. *Proc. R. Soc. Lond. B* **242**, 120 (1990).
9. Baker, R. R. & Bellis, M. A. *Anim. Behav.* **37**, 867–869 (1989).
10. Gage, M. J. G. *Anim. Behav.* **42**, 1036–1037 (1990).

probability and statistics such a rich field.

The literature on many-random-walker processes is sparse, but several interesting results have been derived. For example, in one dimension, a single random walker always returns to its origin, but on the average this takes an infinite amount of time. If, however, three walkers begin at the origin the mean time for one of them to return to the origin is finite<sup>3</sup>. For interacting walkers<sup>4</sup> moving along one dimension that annihilate upon meeting, the probability that  $p$  walkers starting near the origin all survive for  $n$  steps is  $n^{-p(p-1)/4}$ . Such types of problems are intimately involved with interfacial wetting transitions<sup>4</sup>. But note in particular how involving many walkers naturally generates fractional exponents.

Another example concerns the random walk of entangled chains in a polymer melt. The reptation model<sup>5</sup> freezes all polymer chains but one, and calculates the mean first passage time  $T$  for that chain to move away, with snake-like motions, from the chains with which it is entangled. This time  $T$  is found to scale with the chain length  $M$  as  $M^3$ . Experiments, however, yield exponents in the range 3.3–3.4. If one lets all the chains move, some of the entangled neighbours will move in the same direction as the given chain and still be entangled at time  $M^3$ . The many-chain calculation gives  $M^{10/3}$  for the disentangling time<sup>6,7</sup>.

A final example treats many walkers representing mobile defects in a glassy material. The defects can induce a dielectric relaxation upon reaching frozen-in dipoles in the material. For each defect the waiting-time distribution between jumps has such a long tail that the mean jump time is infinite. But when a jump occurs from a finite concentration of defects, not only does the mean time become finite, but the famous stretched exponential law appears as a probability limit distribution which governs the dielectric relaxation<sup>8,9</sup>. The statistics of the minimum jump time from a set are not the same as for a particular member.

The study of random walks has maintained its vitality over three centuries since the pioneering work of Jacob Ber-

noulli. New processes, new questions and new applications ensure a never-ending richness to be mined. The work of Larralde *et al.*<sup>2</sup> opens up a host of further possibilities — of using interacting walkers, of working in fractal spaces, of space-time coupling of trajectory

PALAEOANTHROPOLOGY

## A remote sense for fossils

Bernard Wood

THE latest discoveries<sup>1</sup> of early hominid remains in Ethiopia owe a great deal to the successful application of space-age technology in palaeontology. That technology, remote sensing, holds the promise of helping to deliver further and richer finds both within and beyond East Africa.

The eastern arm of the African Rift valley has furnished much of the evidence for human evolution, sites such as

lengths with velocities (as in turbulence), and so on. These would find applications in fields as diverse as physics and ecology. □

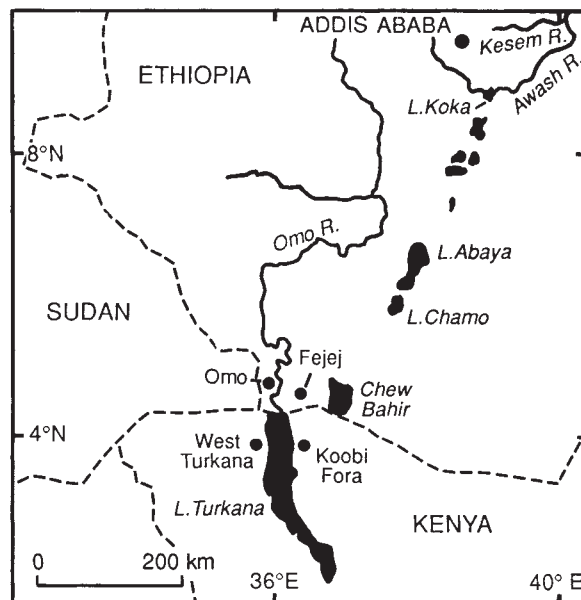
Michael F. Shlesinger is in the Physics Division, Office of Naval Research, Arlington, Virginia 22217-5000, USA.

known to be rich in fossils.

The impasse was broken with the report<sup>2</sup> of the results of a much more systematic survey designed to locate potential fossil sites. From 1988, Berhane Asfaw has been co-ordinating the innovative Palaeoanthropological Inventory of Ethiopia in Addis Ababa. The survey harnesses the powerful techniques of remote sensing. Two modalities, thematic mapping and large-format

camera images, were used by the team, which included members from NASA-Goddard, Los Alamos, the University of California, Berkeley, and the University of Tokyo.

Thematic mapping data are derived from satellite-based radiometers which measure the intensity of reflected sunlight. Whereas earlier Landsat scanners generated spectral data in only four wavebands, with a 76-m resolution and with the range of intensity expressed on a 64-digit scale, the newer equipment provides measures of intensity in six wavebands, with a seventh thematic infrared radiation band, at an increased resolution of 30 m and with an intensity speci-



Sketch map of the principal sites discussed in this article. (Modified from ref. 1.)

Olduvai, Koobi Fora, Omo and Hadar being especially rich sources of the fossilized remains of early hominids. The remains range from those of the earliest known hominid species, *Australopithecus afarensis*, to material which has been allocated to *Homo erectus*. Although research at the sites themselves has become increasingly rigorous, knowledge of their existence has been obtained in a haphazard fashion. Some were pinpointed when explorers and travellers noted finding 'fossil bones', others by extrapolating from regional geological knowledge and looking where sediments 'ought to be'. Field surveys are notoriously hit-and-miss unless controlled by detailed aerial photographs, however, yet the expense of such photographs cannot be justified unless the area is

fied on a 256-digit scale. The Ethiopian project drew upon data enhanced by high-frequency spatial filters and made manifest as composite colour images. The modifications were used to highlight vegetational changes and the occurrence of tephra, and to emphasize fault structures and exposures of sedimentary strata. The large-format camera system images were taken aboard the shuttle *Challenger* in 1984. Substantially larger (23 cm × 46 cm) negatives and more precise optics provided particularly high-resolution photographs which can sustain enlargement to 1:50,000 (the thematic-mapping system cannot practically exceed 1:75,000).

From the data, both major regional geological structures and more local basin topography could be identified.

1. Nossal, R., Haulin, S. & Shlesinger, M. (eds) *J. Stat. Phys.* **65** (5/6), 837–1283 (1991).
2. Larralde, H., Trunfio, P., Havlin, S., Stanley, H. E. & Weiss, G. H. *Nature* **355**, 423–426 (1992).
3. Weiss, G. H. & Rubin, R. J. *Adv. chem. Phys.* **52**, 363–505 (1983).
4. Fisher, M. E. *J. stat. Phys.* **34**, 667–729 (1984).
5. de Gennes, P. G. *J. chem. Phys.* **55**, 572–579 (1971).
6. Scher, H. & Shlesinger, M. F. *J. chem. Phys.* **84**, 5922–5924 (1986).
7. Weiss, G. H., Bendler, J. T. & Shlesinger, M. F. *Macromolecules* **21**, 521–523 (1988).
8. Hamill, W. H. & Funabashi, K. *Phys. Rev.* **B16**, 5523–5527 (1977).
9. Shlesinger, M. F. & Montroll, E. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1280–1283 (1984).



**B destructive**

DIFFERENT functions may have been found for the two mitotic cyclins, A and B, both of which activate p34<sup>cdc2</sup> protein kinase to drive eukaryotic cells through their division cycle. The only clear distinctions between the two have been their kinetics of accumulation and destruction, and their different cellular locations during the cell cycle. F. C. Luca *et al.* (*EMBO J.* **10**, 4311–4320; 1991) now show, in an assay using lysed clam cells, that although truncated versions of both cyclins activate protein kinase activity, only truncated cyclin B can activate the pathway responsible for destruction of full-length, endogenous cyclins. This is not the first claim to have shown functional differences; but the implication of this work *in vitro* is that cyclin B is solely responsible *in vivo* for the activation of cyclin destruction, a process which has to occur before cells can finish mitosis and start the next cell cycle.

**Singular work**

MICROELECTRONICS normally refers simply to very small electronic devices, but might it be better applied to the new circuit of H. Pothier *et al.* (*Europhys. Lett.* **17**, 249–254; 1992) which shunts around individual electrons? Pothier and colleagues first showed that it is possible to handle electrons individually, rather than in countless and undistinguished millions, two years ago when they developed a turnstile that allows electrons to pass through one by one. Their new electron pump, which can move charge in either direction, pulls an electron in at one terminal and ejects it at the other. Like the turnstile, the pump works because once an electron is inside it, its Coulomb repulsion repels any further electrons; unlike the turnstile, the pump (which is just a few micrometres across) can also work with superconducting Cooper pairs of electrons if cooled enough, the authors believe.

**In a twist**

THEORIES to explain the force-generating properties of muscle are rarely as engaging as that put forward by C. Schutt and U. Lindberg (*Proc. natn. Acad. Sci. U. S. A.* **89**, 319–323; 1992). The authors propose that segments of actin filaments contract, pulling on tropomyosin (a helical molecule wrapped round actin) which transmits the summed force to the Z-line of the muscle's sarcomere. Myosin crossbridges (more usually assigned the responsibility of generating force) induce the changes in length of the actin segments and bear the tension of the contracting segments. The evidence for this unconventional scheme originates from the observation of actin 'ribbons', which Schutt and Lindberg believe are the untwisted, extended form of actin.

Such basins occur with a regularity that reflects the regional geology — the survey showed that westward-tilted fault blocks are associated with older sediments exposed on their eastern side, and it predicted that more recent sediments would cover the west-central basin area.

Of particular interest was the identification of pre-Pleistocene sedimentary strata, and the location of further outcrops of the Chorora beds (fossiliferous upper Miocene diatomites) was the key to the team's strategy. The weathered Chorora exposures have a characteristic thematic mapping reflectance pattern, which allowed the scientists to locate hitherto unidentified exposures of Chorora beds and led them to concentrate on the Kesem–Kabena basin (top right of the map). The basin is north of Lake Koka in the Awash river system, towards the Afar depression, and is upstream of the well-known sites of Hadar and the Middle Awash.

The next step was just that — to survey the Kesem–Kabena area by vehicle and on foot. From the remote sensing information it seemed that Pliocene–Pleistocene strata should be exposed along the fault bordering the northwestern limits of the basin, and this is where the search was concentrated. In the event, several sites were located by ground-based transects. In one locality (K-K1), Pliocene vertebrate fossils are abundant and the fossiliferous sediments are likely to be older than 2.3 million years (Myr). Two others (K-K6 and K-K4) are more recent (1.0 Myr) and include rich assemblages of Acheulean artefacts as well as vertebrate fossils<sup>3</sup>.

**Early hominid teeth**

As well as surveys in the Awash river system, the inventory project's 1989 field season took in the south of Ethiopia; here the potential of the Fejej area became clear<sup>4</sup>. The tuff-correlation studies of Brown and his colleagues<sup>5,6</sup> had already shown that the Omo, Koobi Fora and West Turkana fossil sites all sample sediments that are derived from a single lake basin and its associated river systems.

The Fejej group of sites is located in the last relatively unexplored portion of the basin, its northeast quadrant. In 1990 the area was prospected by a team with members from the State University of New York at Stony Brook, University of California, Los Angeles, and the US Geological Survey. Among the fossils recovered by Fleagle and colleagues<sup>1</sup> from below the early Pliocene Harr Basalt at site FJ-4 are specimens of early hominid teeth. They clearly resemble the dental remains of *A. afarensis* found at Hadar<sup>7</sup> and Laetoli<sup>8</sup>. Their morphology is unsurprising, but they confirm the

presence of *A. afarensis* in the Omo regional deposits and the <sup>40</sup>Ar/<sup>39</sup>Ar dates from an overlying basalt suggest that the Fejej hominid remains are at least as old as those from Laetoli (3.6–3.7 Myr) and maybe up to 0.7 Myr older.

**Moratorium lifted**

Doubtless encouraged by the success of the inventory project, in 1990 the Ethiopian authorities lifted the eight-year moratorium on palaeoanthropological field research. Two research groups returned to familiar territory, the team from the Institute of Human Origins (Johanson, Kumbel *et al.*) to the Hadar site and the University of California, Berkeley team (Clark, White *et al.*) to the Middle Awash. Both have come up with further palaeontological and archaeological evidence for early occupation. The 15 new hominids recovered from Hadar include cranial, mandibular and postcranial remains.

Hominid palaeontologists are eagerly awaiting details of this new evidence, for it will be the first real test of the 'single-species' hypothesis as it has come to be applied to the Pliocene hominids which have been recovered from Ethiopia, Tanzania and elsewhere in East Africa. Most, but not all, recent studies of the Hadar and Laetoli remains have concluded that the collections sample a single, highly, but not excessively, dimorphic early hominid<sup>9,10</sup>. Such a hypothesis implies that both the degree and pattern of variation within *A. afarensis* should be consistent with that predicted by a combination of comparative and phylogenetic analysis. At present we are better able to predict what this means for the cranium, mandible and the dentition<sup>11</sup> than for the postcranial skeleton. But studies in progress should provide the necessary background data, and we shall soon be equipped to make a precise expression of the probability that one, or more, species are represented within the Ethiopian evidence. □

Bernard Wood is in the Department of Human Anatomy and Cell Biology, University of Liverpool, Liverpool L69 3BX, UK.

1. Fleagle, J. G. *et al.* *J. hum. Evol.* **21**, 145–152 (1991).
2. Asfaw, B. *et al.* *Nat. geog. Res.* **6**, 418–434 (1990).
3. Wolde Gabriel, G. *et al.* *J. Field Archaeol.* (in the press).
4. Asfaw, B. *et al.* *J. hum. Evol.* **21**, 137–143 (1991).
5. Cerling, T. E. *et al.* *Nature* **279**, 118–121 (1979).
6. Brown, F. H. *et al.* in *Ancestors: The Hard Evidence* (ed. Delson, E.) 82–90 (Liss, New York, 1985).
7. Johanson, D. C. *et al.* *Am. J. phys. Anthropol.* **57**, 545–603 (1982).
8. White, T. D. *et al.* in *Ancestors: The Hard Evidence* (ed. Delson, E.) 138–152 (Liss, New York, 1985).
9. Kimbel, W. H. *et al.* in *Ancestors: The Hard Evidence* (ed. Delson, E.) 120–137 (Liss, New York, 1985).
10. Wood, B. A. *Koobi Fora Research Project, Vol. 4, Hominid Cranial Remains* (Clarendon Press, Oxford, 1991).
11. Wood, B. A. *et al.* *J. Anat.* **174**, 185–205 (1991).

# Is no SNUs good news?

Lawrence M. Krauss

ANY hopes that the long-standing 'solar neutrino' problem might go away once new neutrino detectors came on line appear to have faded with the first results from the Soviet-American Gallium Experiment (SAGE), reported in *Physical Review Letters* last month<sup>1</sup>. The authors report a deficit in the number of neutrinos coming from the Sun, compared with the number expected to be produced by the main solar nuclear-energy-producing reactions. This result supports previous evidence that something is wrong with our understanding of either solar or neutrino physics, and further points towards the latter.

The neutrinos that are being sought are emitted directly from the solar core, where they are produced in nuclear reactions which fuel the Sun. Because the light we see is emitted from the solar surface, it is only by detecting neutrinos that we can probe directly the physical processes associated with solar energy production, and thus solar evolution. The SAGE experiment is located in a deep underground laboratory at the Baksan Neutrino Observatory in the north Caucasus. It involves detecting the production of about one atom of  $^{71}\text{Ge}$  per day in a detector containing 30 tons of liquid  $^{71}\text{Ga}$ , produced by interactions of electron neutrinos ( $\nu_e$ ) originating in fusion reactions which convert hydrogen to helium.

To date, there have been two other solar neutrino experiments — a radiochemical experiment involving the production of  $^{37}\text{Ar}$  from  $^{37}\text{Cl}$  in an underground tank at the Homestake gold mine in the United States<sup>2</sup>, which has been taking data for over 20 years; and a real-time detector of neutrino events in a large underground water tank at the Kamioka facility in Japan<sup>3</sup>, which has been reporting results for three years. These have both consistently reported neutrino event rates substantially below those predicted by standard solar models, leading to the solar neutrino problem. The solution could require dramatic new physics associated with either astrophysical processes deep in the solar interior, or the properties of neutrinos themselves. In either case, the consequences could be profound.

The crucial feature that differentiates the  $^{71}\text{Ga}$  experiment from the earlier solar neutrino experiments is the fact that the latter are primarily sensitive to only the highest-energy component of the solar neutrino flux, one associated with a rather minor side reaction ( $^8\text{B} \rightarrow ^8\text{Be} + e^- + \nu_e$ ) in the nuclear chain that causes four hydrogen nuclei to fuse and

make the nucleus of a helium atom. On the other hand, the threshold (0.2332 MeV) for the reaction  $\nu_e + ^{71}\text{Ga} \rightarrow ^{71}\text{Ge} + e^-$  is extremely low, allowing the  $^{71}\text{Ga}$  experiment to be primarily sensitive to the dominant component of the neutrino flux from the Sun. This component is associated with the first step of the fusion reaction chain ( $^1\text{H} + ^1\text{H} \rightarrow ^2\text{H} + e^+ + \nu_e$ ), and so is directly linked to the solar luminosity, a well measured quantity. Hence, any suppression of this neutrino signal would give unambiguous evidence that the solar neutrino problem originates in the fundamental physics of neutrinos. But were the observed signal to agree with that predicted, it might suggest that the solar neutrino problem is restricted to the high-energy boron nuclear reaction network, or perhaps that the other detectors have been giving

spurious results.

For these reasons, the SAGE results have been eagerly awaited. The data have been a long time in the making. The basic chemistry associated with the  $^{71}\text{Ga}$  experiment was first explored 14 years ago<sup>4</sup>. SAGE uses four Teflon-lined chemical reactors, each holding about 7 tons of liquid gallium. For each measurement a small amount of natural germanium carrier is added to each reactor; after 3–4 weeks this carrier and any  $^{71}\text{Ge}$  atoms produced are chemically extracted as  $\text{GeCl}_4$ , after the addition of hydrochloric acid and hydrogen peroxide, and this is synthesized and purified into germane gas ( $\text{GeH}_4$ ). The overall extraction efficiency of germanium by this procedure is about 80 per cent. The germane gas is mixed with xenon, and then put into low-background gas proportional counters. These in turn are placed in shields, where they are sensitive to the X-rays and electrons produced when  $^{71}\text{Ge}$  atoms radioactively decay, and monitored for 2–3 months.

## Black hole is M87's bright spot

WHEN they reported in 1978 the possible discovery of a black hole in the elliptical galaxy M87, W. L. W. Sargent *et al.* suggested that the projected Space Telescope would provide the extra spatial resolution needed to confirm their claim. Now, 14 years later, the much maligned Hubble Space Telescope has done just that. T. R. Lauer, S. M. Faber and colleagues claimed at the recent meeting of the American Astronomical Society (Atlanta, 13–17 January 1992). The evidence comes from the brightening of the galactic light near the nucleus. This same brightening was seen in the earlier observations of Sargent *et al.*, who attributed it to the increasing density of stars in that direction. Comparing the stellar distribution with that in a normal galaxy, those authors concluded that the anomalously concentrated packing of stars, and their velocities (revealed spectroscopically), could be accounted for only if there was a central concentration, probably a black hole, comprising 5 billion ( $10^9$ ) solar masses in the nucleus. The new high-resolution image shown here, taken with the Hubble's Wide Field Camera, confirms that picture, and the astronomers conclude that the black hole is equivalent to 2.6 billion solar masses. Some of the reasons why M87 was chosen in the search for a massive black hole are also apparent in the figure. The galaxy is the archetypal radio or active galaxy. At its centre is a brilliant source of radio waves, shown here also to be the source of intense (non-stellar) optical emission. Also it ejects one of the most dramatic astrophysical jets, stretching 5,000 light years to the right on this picture. This, and possibly its counterpart, illuminate two giant radio lobes that extend beyond the visible extent of the galaxy. All this activity is thought to result from a disk of dense plasma in the galactic core, heated as it spirals onto the central black hole. The Hubble scientists promise further spectroscopic studies to strengthen their conclusion. Another reason for looking at M87 is that it is relatively close, so affording us a good view. Will Hubble now be turning its cameras towards NGC6240, which is even closer and may contain, it was suggested last year, a black hole ten times more massive?



R. P.



The published data appear not only to confirm the solar neutrino problem, but, taken at face value, imply that the source of the problem lies with the neutrinos. The explicit result quoted is as follows: using units developed for the  $^{37}\text{Cl}$  experiment, called solar neutrino units (SNU), in which 1 SNU =  $10^{-36}$  events per target atom per second, the SAGE collaboration reports, after five uncontaminated observing runs, an interaction rate of  $20^{+15}_{-20} \pm 32$  SNU — about three germanium atom decays observed during the course of the experiment. (The uncertainties are statistical and systematic, respectively.)

From this the researchers claim an upper limit of 79 SNU at the 90 per cent confidence level, which is to be compared with rates of  $132^{+20}_{-17}$  or  $125 \pm 5$  SNU (at 99 and 67 per cent confidence respectively) predicted by various standard solar models<sup>5,6</sup>. The upper limit of 79 SNU is particularly intriguing, because it has been argued that this is also precisely the theoretical lower limit on the  $^{71}\text{Ga}$  event rate, assuming only that the Sun radiates energy as fast as it generates it (so it is in equilibrium), and that the associated neutrinos are unaffected as they make their way to the Earth. Thus, an observation of a rate less than 79 SNU would lay the burden of the solar neutrino problem squarely on the neutrinos themselves. This would provide the first empirical evidence that the standard model of elementary particle physics is incomplete.

What neutrino physics could suppress the flux of electron neutrinos from the Sun? The SAGE result alone cannot easily distinguish between the possibilities so far suggested. The most investigated proposal has to do with the possibility of a mass difference between the electron neutrino and its cousins, the muon and tau neutrinos which are invisible to gallium detectors. If neutrinos have a mass, their mass eigenstates need not be identical to the eigenstates of the weak interaction that creates them, so that as they traverse the solar interior or on their journey to the Earth, electron neutrinos could oscillate into invisible muon or tau types. A mass difference 100 million times smaller than the electron's mass is all that is needed. Masses of this magnitude have been predicted if the three known forces are unified into a single grand unified theory at cosmological energy.

Nonetheless the SAGE result is not yet definitive. This is the first set of results obtained using this technique, and it is based upon only the first five runs, during which time the cumulative total number of events observed by counting events for two  $^{71}\text{Ge}$  half-lives after each run was about nine, including backgrounds. Of this total, only two

were assigned as signal, compared with 13 predicted by standard solar models. A single extra atom per run in the signal would change the 90 per cent confidence limit by 35 SNU. More important, the experiment has yet to be calibrated with a neutrino source. This is essential to guarantee that the efficiency for detecting  $^{71}\text{Ge}$  atoms produced by solar neutrinos is the same as for the natural germanium carrier. In addition it should be noted that the source of the background, which dominates the data and is itself almost as large as the predicted solar signal, is not yet clearly determined. And there are still puzzling experimental data which are not addressed by the SAGE result. In particular, there is an apparent anti-correlation of the event rate in the Homestake  $^{37}\text{Cl}$  experiment with the solar cycle which, while statistically significant, is very difficult to understand theoretically. Also, the observation of a 17-kiloelectronvolt neutrino state (see Nelson's News and Views article<sup>7</sup>) would, if correct, make it much more difficult to assign the light masses to two neutrino species necessary for the neutrino oscillations.

Another gallium experiment, the GALLEX experiment, run by a European-US-Israeli collaboration and located in the Gran Sasso underground laboratory in Italy, is currently on line using 30 tons of gallium in the form of gallium chloride. After several frustrating years overcoming and understanding quite complex chemical contamination problems, the researchers have been taking data for five months, and have stated that they will report results after amassing ten months of data — enough to make statistically unambiguous statements. It is prudent to await these GALLEX data, which should be available before a calibration of the SAGE detector is completed, before one proclaims the dawn of a new era. Nevertheless, if GALLEX agrees with SAGE, solar neutrinos will have opened a new window on fundamental physics, and the race will then be on to develop a new generation of solar neutrino detectors capable of exploring the details. □

Lawrence M. Krauss is in the Center for Theoretical Physics, Yale University, PO Box 6666 Yale Station, New Haven, Connecticut 06511, USA.

1. Abazov, A. I. et al. *Phys. Rev. Lett.* **67**, 3332–3335 (1991).
2. Davis, R. et al. in *Proc. Thirteenth Int. Conf. Neutrino Phys. Astrophys.* (eds Schneps, J. et al.) 518 (World Scientific, Singapore, 1989).
3. Hirata, K. D. et al. *Phys. Rev. Lett.* **65**, 1297–1300 (1990).
4. Bahcall, J. N. *Neutrino Astrophysics* (Cambridge University Press, Cambridge, 1989).
5. Bahcall, J. N. & Ulrich, R. *Rev. Mod. Phys.* **60**, 297–372 (1988).
6. Turck-Chièze, S., Cahen, S., Casse, M., Doom, C. *Astrophys. J.* **335**, 415–424 (1988).
7. Nelson, H. N. *Nature* **350**, 461–462 (1991).

## Counting stars

THE Hubble space telescope was planned to have the highest possible resolution over a narrow field of view. Daedalus's new space-telescope project has the converse ambition. His 'Skygazer' will combine extremely bad resolution with the ultimate field of view. It will look at all the sky all the time.

Ever since Olbers propounded his famous paradox, astronomers and cosmologists have argued about the intrinsic brightness of the night sky. In the visible waveband, an average square degree of sky delivers about 20 femtowatts to a detector 1 centimetre across: perhaps 60,000 photons per second. Such a modest photon flux could easily be counted directly by a photomultiplier or an avalanche photodiode array. A spacecraft carrying forty thousand such detectors in a spherical fly's-eye arrangement would intercept about 2.5 billion photons a second from the whole celestial sphere, and could count them all. Its photometric accuracy would be absolute.

The Skygazer will have to be shielded from the Sun, which would blind any photodetector that glimpsed it (the Hubble telescope has the same problem). It may also need to be screened from the Earth and Moon, and should be launched into the most distant orbit consistent with adequate telemetric and command links. But it will then provide the most accurate sky-brightness information any cosmologist could wish for.

It will also act as a vigilant watchdog for new celestial developments of all kinds. Let a nova start anywhere in the Galaxy, and the Skygazer will spot the rising photon count. Its direction will be that of the photodetector whose count is rising; an Earth-based telescope could then search that region of sky. Variable stars, stellar collisions, flares and many other fascinating astronomical phenomena could be spotted early enough to follow their full evolution.

Some bold cosmological speculations could also be tested. Daedalus would like to find a correlation between the individual photon counts from opposite points of the sky. This would prove that celestial antipodes are the same, so that their light reaches us from two opposite paths round the whole Universe. Similarly, photon correlation between widely separated astronomical objects would show them to be different views of the same thing, doubly imaged by some outrageous gravitational-lens effect. And if the summed sky count is found to be slowly declining year by year, the steady-state theory of the Universe can finally be abandoned. Cosmic darkness will be closing in, and the date of Doomsday will be known.

DAVID JONES

# Doping of C<sub>60</sub> with iodine

SIR — The announcement by H. Sekine *et al.* of superconductivity in I- and Br-doped fullerenes generated much curiosity (see ref. 1) because iodine-fullerene mixing was not expected, based on comparison with graphite-intercalated compounds (for example, ref. 2), and because superconductivity would reflect hole doping rather than electron doping. Hole doping is particularly intriguing, given the ease with which metallic and superconducting electron-doped fullerenes can be prepared<sup>3</sup>.

We have used X-ray photoelectron spectroscopy to examine I mixing with C<sub>60</sub> and the subsequent energy-level spectrum. An ultrahigh vacuum I<sub>2</sub> dosing source was constructed, based on the electrochemical dissociation of AgI (ref. 4). Advantages of such a source are the precise control over dosing and minimal exposure of the chamber to the reactive gas. Calibration of the source was based on the saturation coverage of a cleaved GaAs(110) surface, which for Cl<sub>2</sub> is about half of a monolayer, or  $4.4 \times 10^{14}$  atoms cm<sup>2</sup> (ref. 5). A 75-Å C<sub>60</sub> layer was formed on GaAs and annealed to about 150 °C to form a well-ordered layer, as shown by scanning tunnelling microscopy<sup>6</sup>.

Initial exposures of clean fullerene films to about  $10^{15}$  I<sub>2</sub> molecules cm<sup>2</sup> at 100 °C showed only trace amounts of I, and analysis of the attenuation of the

GaAs and C<sub>60</sub> emission demonstrated that a small amount was intermixed with the film, with a composition of approximately C<sub>60</sub>I<sub>0.04</sub>. The effect of I incorporation was to shift the C 1s emission 0.2 eV to lower binding energy, but the spectral signatures of the fullerene showed no changes that would indicate compound formation. Instead, the changes were analogous (but in the opposite directions) to those observed following the addition of small amounts of K to C<sub>60</sub> layers, an effect related to the pinning of the Fermi level near the edge of the LUMO-derived conduction band and screening by these electrons of the core holes<sup>7</sup>. Hence, I doping induces Fermi level movement towards valence band states derived from HOMO. Depletion of valence states by acceptors should result in a dramatic decrease in the C 1s binding energy, but this was not observed, suggesting that complete charge transfer to I does not occur. The binding energy of the I 3d core level was 0.1 eV lower than both I<sub>2</sub> dissociated on the GaAs(110) surface and CdI<sub>2</sub> layers with the same linewidth, suggesting that I was atomic and at least partially ionic.

Higher exposures to I<sub>2</sub> did not increase the amount of I incorporation, in contrast to the behaviour for alkali doping, where mixing led to fulleride formation<sup>3,7</sup>. To increase the amount of I in the film, we decreased the substrate temperature and illuminated the surface during I<sub>2</sub> exposure to create excited state I<sub>2</sub>. The fullerene films were exposed to  $2 \times 10^{16}$  I<sub>2</sub> molecule cm<sup>2</sup>, providing about 40 I atoms for each fullerene if the sticking coefficient were unity. Because the iodine signal never increases above the level formed by the initial exposure, C<sub>60</sub>I<sub>0.04</sub>, the formation of I-containing compounds from the vapour phase is highly unfavourable.

As an alternative method of providing I for mixing with C<sub>60</sub>, we deposited C<sub>60</sub> molecules onto an I-saturated GaAs surface. I was not released from the GaAs bonds into the fullerene overlayer because I mixing was not energetically favourable. Similarly, mixing was not promoted either by illumination (which would provide electrons and holes at the surface and could enhance surface chemistry) or by heating to any temperature below that needed to desorb the fullerenes.

A difficulty inherent with iodine doping from the AgI source is that the cell generates I<sub>2</sub> molecules rather than I atoms. The 1.33-Å atomic radius of I is sufficiently small that diffusion in the interstices could occur after dissociation of the larger molecule. To enhance I<sub>2</sub> dissociation before adsorption, a hot W

filament was placed in line-of-sight with the I<sub>2</sub> generator. In this case, exposure to about  $2 \times 10^{16}$  I<sub>2</sub> molecule cm<sup>2</sup> produced an intermixed overlayer with a stoichiometry of C<sub>60</sub>I<sub>0.3</sub>. The C 1s emission was shifted rigidly to lower binding energy relative to the pure C<sub>60</sub> layer by the same amount as was observed for the C<sub>60</sub>I<sub>0.04</sub> layer formed without the filament. Much higher exposures failed to yield a measurable increase of I-C<sub>60</sub> mixing.

In summary, the exposure of fullerene layers to I from the vapour phase results in slight doping but not the formation of a distinct compound. It remains to be seen if stable phases exist, but our results suggest that electron transfer from the HOMO of C<sub>60</sub> is unlikely.

T. R. OHNO  
G. H. KROLL  
J. H. WEAVER

Department of Materials Science and  
Chemical Engineering,  
University of Minnesota,  
Minneapolis,  
Minnesota 55455, USA

L. P. F. CHIBANTE  
R. E. SMALLEY

Rice Quantum Institute and  
Departments of Chemistry and Physics,  
Rice University, Houston,  
Texas 77251, USA

1. Swinbanks, D. *Nature* **352**, 749 (1991).
2. Hérold, A. in *Chemical Physics of Intercalation* (eds Legrand, A. P. & Flandrois, S.) 3–45 (Plenum Press, New York, 1987).
3. Haddon, R. C. *et al.*, *Nature* **350**, 320 (1991).
4. Mowbray, A. & Jones, R. G. *J. Vac. Sci. Technol.* **A7**, 3373 (1989).
5. DeLouise, L. A. *J. Vac. Sci. Technol.* **A9**, 1732 (1991).
6. Li, Y. Z. *et al. Science* **253**, 429 (1991).
7. Benning, P. J. *et al. Phys. Rev. B* (in the press).

## Kiwi's egg

SIR — Speculation has recently been revived (see S. J. Gould, *Bully for Brontosaurus*, Hutchinson Radius, 1991) on the issue of the kiwi, and how natural selection might account for its relatively enormous egg, which is of no obvious advantage to the chick and a considerable disadvantage to the hen. I suggest that we find the problem baffling only because, by the time we began wondering about it, the moa had become extinct.

Before man arrived in New Zealand, the kiwi had no mammalian predators. But large omnivorous ratites like the moa gobble up eggs laid on the ground, except those in the large clutches deposited by their own harems. A kiwi's egg added to such a clutch might escape their predation, and the bigger it was the more easily it would avoid detection and eviction. Later, when the moa vanished and the rat arrived, kiwi survivors would be those which most rapidly learned to scratch holes in the ground.

ELAINE MORGAN

24 Aberffwd Road,  
Mountain Ash,  
Mid Glamorgan CF45 4AR, UK

## Mount Etna CO<sub>2</sub> may affect climate

SIR — Allard *et al.*<sup>1</sup> recently estimated current summit-crater emissions from Mount Etna of  $13 \pm 3$  Tg yr<sup>-1</sup> CO<sub>2</sub>, with a similar contribution to the atmosphere from diffuse flank emissions. Surprisingly, this yields a total CO<sub>2</sub> flux from Etna of about  $0.6 \times 10^{12}$  mol yr<sup>-1</sup>, or about 10% of the estimated global CO<sub>2</sub> flux from all metamorphic and mantle sources (about  $6 \times 10^{12}$  mol CO<sub>2</sub> yr<sup>-1</sup>, ref. 2). Etna is known to have been continuously active since at least the fourth century BC<sup>3</sup>. If Mount Etna's CO<sub>2</sub> emissions were continuous and sustained for much of its long history, then the use of a carbon-cycle model<sup>4</sup> using Berner's<sup>5</sup> non-linear weathering-rate formulation indicates that CO<sub>2</sub> fluxes from Mount Etna alone may have been responsible for about 60 of the 280 parts per million pre-industrial CO<sub>2</sub> concentration.

This suggestion implies that variability in Mount Etna's CO<sub>2</sub> emissions could have climatic consequences, producing



atmospheric  $\text{pCO}_2$  variations of about two-thirds the magnitude of the about 90 p.p.m. glacial-interglacial  $\text{pCO}_2$  difference<sup>6</sup>. Models to explain these glacial-interglacial variations in atmospheric  $\text{CO}_2$  have centred on changes in ocean circulation<sup>7</sup>, but no entirely satisfactory explanation has yet been proposed. Recent evidence that volcanic eruptions may follow a ~23,000-yr cycle<sup>8</sup> (perhaps as a result of changes in sea level)<sup>9</sup> suggests possible orbital pace-makers for volcanic  $\text{CO}_2$  emissions. Could a few volcanoes such as Etna significantly affect the atmospheric  $\text{CO}_2$  concentration on the  $10^4$ – $10^5$ -yr timescale?

KEN CALDEIRA

Earth System Science Center and  
Department of Geosciences,  
The Pennsylvania State University,  
University Park,  
Pennsylvania 16802, USA

MICHAEL R. RAMPINO

Earth Systems Group,  
Department of Applied Science,  
New York University,  
New York, New York 10003, USA

1. Allard, P. *et al.* *Nature* **351**, 387–391 (1991).
2. Berner, R. A., Lasaga, A. C. & Garrels, R. M. *Am. J. Sci.* **283**, 641–683 (1983).
3. Chester, D. K., Duncan, E. M., Guest, J. R. & Kilburn, C. R. *J. Mount Etna: the Anatomy of a Volcano* (Stanford University Press, 1985).
4. Caldeira, K. & Rampino, M. R. *Geophys. Res. Lett.* **15**, 987–990 (1991).
5. Berner, R. A. *Am. J. Sci.* **291**, 339–376 (1991).
6. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
7. Broecker, W. S. & Denton, G. *Quart. Sci. Rev.* **9**, 305–341 (1990).
8. Paternite, M., Labeyrie, J., Guichard, F., Mazaud, A. & Maitre, F. *Earth, planet. Sci. Lett.* **98**, 166–174 (1990).
9. Rampino, M. R., Self, S. & Fairbridge, R. W. *Science* **206**, 826–829 (1979).

## More questions on forest decline

SIR — Ross and De Serves<sup>1</sup> criticize the recent proposal by Becker *et al.*<sup>2</sup> that  $\text{H}_2\text{O}_2$  formation via the reaction of terpenes with ozone in the presence of water could make a significant contribution to the mixed-layer  $\text{H}_2\text{O}_2$  budget, thereby contributing to forest decline. We agree that the yields reported by Becker *et al.* are too low to make a significant contribution to the  $\text{H}_2\text{O}_2$  budget of the mixed layer compared to  $\text{H}_2\text{O}_2$  formation via peroxy radical disproportionation.

Becker *et al.* in their reply<sup>3</sup> to the criticism by Ross and De Serves suggest that the true yields may be much larger than their reported values because, for the high concentrations of reactant used by them, the reaction of the Criegee biradicals with aldehydes produced in the reaction would compete with their reaction with water. But the independence of the  $\text{H}_2\text{O}_2$  yields over a wide reactant concentrations range, as re-

ported by Becker *et al.*, implies that such a competition cannot be important under their conditions. Our recently reported significantly higher yields for the formation of  $\text{H}_2\text{O}_2$  via the new process<sup>4</sup> suggests that under some conditions that contribution of this process to the mixed layer  $\text{H}_2\text{O}_2$  budget may not be negligible. Using fluxes for terpenes quoted by Ross and De Serves, for example, and assuming that about 25% of the  $\beta$ -pinene and about 50% of the other terpenes are oxidized by reaction with ozone<sup>5</sup>, the rate of  $\text{H}_2\text{O}_2$  formation is estimated to be about 52 parts per  $10^{12}$  by volume per hour (p.p.t.v.  $\text{h}^{-1}$ ) in the mixed layer. This can be compared to 95 p.p.t.v.  $\text{h}^{-1}$  calculated for the disproportionation process for a  $\text{HO}_2$  concentration of 20 p.p.t.v. (ref. 6).

Becker *et al.*<sup>3</sup> suggest that our observations of high  $\text{H}_2\text{O}_2$  yields are due to interference from ozone in the wet chemical system used for peroxide analysis. There is no basis for this conclusion. Previous studies have shown that interference by ozone using the

fluorimetric method for  $\text{H}_2\text{O}_2$  determination is very small<sup>7</sup>. This was confirmed in our experiments. Introduction of 150 parts per  $10^9$  ozone into the reaction chamber in the absence of alkene initially indicated the formation of about 70 p.p.t.v. of  $\text{H}_2\text{O}_2$  which rapidly decreased to about 30 p.p.t.v. This corresponds to less than 1% of the  $\text{H}_2\text{O}_2$  formed in the experiments in the presence of alkene.

R. SIMONAITIS

K. J. OLSZYNA

J. F. MEAGHER

Atmospheric Science Department,  
Tennessee Valley Authority,  
Muscle Shoals,  
Alabama 35660, USA

1. Ross, H. B. & De Serves, C. *Nature* **352**, 672 (1991).
2. Becker, K. H., Brockman, K. J. & Bechara, J. *Nature* **346**, 257–258 (1990).
3. Becker, K. H., Brockman, K. J. & Bechara, J. *Nature* **352**, 672–673 (1991).
4. Simonaitis R., Olszyna, K. J. & Meagher, J. F. *Geophys. Res. Lett.* **18**, 9–12 (1991).
5. Atkinson, R., Hasegawa, D. & Ashmann, S. M. *Int. J. Chem. Kinetics* **22**, 871–887 (1990).
6. Trainer, M. *et al.* *J. geophys. Res.* **92**, 11879–11894 (1987).
7. Lazrus, A. L. *et al.* *Analyt. Chem.* **58**, 594–597 (1986).

## Maintenance of MHC polymorphism

SIR — Although the finding of Hill *et al.*<sup>1</sup>, that specific human class I and class II MHC alleles are associated with resistance to *Plasmodium falciparum* malaria, is of undoubted importance, their interpretation of the evolutionary mechanisms underlying their findings is questionable. Hill *et al.* give the impression that their data may favour the hypothesis of frequency-dependent rather than overdominant selection as the mechanism of maintenance of MHC polymorphisms. In fact, their data are not consistent with the former hypothesis, and their study does not constitute a true test of the latter.

The overdominance hypothesis<sup>2,3</sup> depends on the exposure of the population to two or more pathogens (either separate species or antigenically distinct strains of the same species) to which different MHC alleles confer resistance. We previously wrote: "As a particular class I antigen may provide enhanced recognition of a particular pathogen, individuals expressing various types of class I antigen may have a selective advantage in a population exposed to a diverse array of pathogens; such individuals will be those which are heterozygous at most or all loci"<sup>3</sup>. On this hypothesis, one would expect to find (as Hill *et al.* report in the case of HLA-B\*53) that heterozygotes for an allele conferring resistance to a single pathogen are no more resistant to that pathogen than are homozygotes for the same allele. In fact, *P. falciparum* populations are polymorphic with respect to proteins that elicit an immune

response<sup>4</sup>, and it is conceivable that different MHC alleles confer resistance to different strains of this parasite. But because Hill *et al.* did not consider polymorphism in *P. falciparum* and did not study other pathogen species, their results are not directly relevant to the overdominance hypothesis.

In the case of class II MHC haplotypes, Hill *et al.* present evidence that, among individuals having the protective DRw13.02 haplotype, there is a higher proportion of heterozygotes among those with severe malarial anaemia than among mild controls. This suggests that the fitness of DRw13.02 homozygotes is higher than that of DRw13.02 heterozygotes. If this is true, it has nothing to do with the maintenance of MHC polymorphism. In the absence of other factors, such as the presence of other pathogens to which other haplotypes confer resistance, these apparent fitness values amount to an absolute advantage for the DRw13.02 haplotype. Therefore, if no other factors are involved, the result should eventually be fixation of the DRw13.02 haplotype.

Furthermore, it is worth noting that the data of Hill *et al.* provide no support for the hypothesis of frequency-dependent selection. The only model of frequency-dependent selection that is as effective as that of overdominant selection for maintaining the type of multi-allelic, long-term polymorphism characteristic of the MHC is a model of rare allele advantage<sup>5</sup>. But the data of Hill *et al.* for both class I and class II show that the protective alleles have high frequen-

cies and that rare alleles are not associated with resistance to malaria.

AUSTIN L. HUGHES  
MASATOSHI NEI

Department of Biology and Institute  
of Molecular Evolutionary Genetics,  
The Pennsylvania State University,  
University Park,  
Pennsylvania 16802, USA

HILL ET AL. REPLY — Our study showed that certain HLA alleles are significantly associated with protection from severe malaria. Hughes and Nei do not question our inference that these alleles appear to have risen in frequency through natural selection by malaria, nor our conclusion that the data provide important support for the theory of parasite-driven HLA polymorphism. Instead, they address the issue of the type of selection involved, an interesting and controversial question on which we did not commit ourselves.

It is now widely believed that HLA polymorphism results from some balancing selection pressures and, in theory, diversity may be maintained by overdominant, frequency-dependent or fluctuating selection or, perhaps most likely, some combination of all three<sup>6</sup>. Our study was not designed to distinguish between these but, given the shortage of other field data in humans, the Gambian results are of interest.

The analysis of overall heterozygosity rates is statistically more powerful than comparisons of risks associated with heterozygosity or homozygosity for particular alleles. Heterozygotes at certain HLA loci were not at decreased risk of severe malaria compared to homozygotes<sup>1</sup>, but Hughes and Nei deny that this is a test of the overdominance hypothesis because only one species of pathogen was studied. However, because *Plasmodium falciparum* shows striking polymorphism in many of its immunodominant antigens<sup>4</sup> and most Gambian children with malaria harbour more than one antigenically distinct parasite strain<sup>7</sup>, it was certainly possible that HLA heterozygotes, who could recognize a greater range of epitopes in these parasite strains, might show greater resistance to severe malaria. For HLA-B and DR-DQ genes, the loci for which we identified resistance alleles, the observed changes in overall heterozygosity were actually in the opposite direction. From the point of view of an HLA molecule, it does not matter whether different foreign peptides come from distinct species or distinct strains.

The minority (rare allele) advantage type of frequency-dependent selection is as good as symmetrical overdominant models at maintaining allelic diversity<sup>5</sup>, and the finding that alleles with frequencies of 0.13–0.15 are protective is not

incompatible with minority advantage models. The cardinal feature of the latter is that the fitness of an allele (or genotype) should decrease as its frequency increases. These West African HLA alleles may have been even more protective when they were less common. Although the simple minority advantage model of Takahata and Nei<sup>5</sup> assumed identical genotypic fitnesses for all alleles at a given frequency this seems biologically unrealistic. It would be useful to examine the effect of (fluctuating) allelic differences in frequency-dependent fitness on the preservation of allelic diversity: increasing biological plausibility in overdominant models by allowing asymmetric overdominance leads to some allele loss<sup>5</sup>.

The higher fitness of HLA-DRw13.02 homozygotes than heterozygotes against severe malaria, suggested by our study<sup>1</sup>, is compatible with a minority advantage type model but not with simple overdominance: in the former the allele does not reach fixation because its advantage diminishes and then disappears as it becomes more frequent. Moreover, it was not shown that rare alleles are not associated with resistance to malaria. In case-control studies of reasonable size only fairly common alleles can be adequately assessed: rare alleles are too infrequent to show statistically significant frequency differences, particularly in the direction of a protective effect, unless the associated relative risks are improbably large.

Although the Gambian data may be

seen by some to support frequency-dependent selection, our remarks were phrased to contrast with Hughes and Nei's earlier paper entitled "Pattern of nucleotide substitution at a major histocompatibility complex loci reveals overdominant selection"<sup>3</sup>. This title was inaccurate, in that the data analysed provide evidence of selection that may take various forms and need not be overdominant. Carefully designed field studies to address this issue in humans are needed, and are increasingly practicable.

ADRIAN V. S. HILL  
DOMINIC KWIATKOWSKI  
ANDREW J. MCMICHAEL

Institute of Molecular Medicine,  
University of Oxford,  
John Radcliffe Hospital,  
Oxford OX3 9DU, UK

BRIAN M GREENWOOD

MRC Laboratories,  
Fajara, The Gambia

STEPHEN BENNETT

Department of Epidemiology and  
Population Sciences,  
London School of Hygiene and  
Tropical Medicine,  
London WC1E 7HT, UK

1. Hill, A. V. S. et al. *Nature* **352**, 595–600 (1991).
2. Doherty, P. C. & Zinkernagel, R. M. *Nature* **256**, 50–52 (1975).
3. Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
4. Lockyer, M. J., Marsh, K. & Newbold, C. I. *Molec. biochem. Parasit.* **37**, 275–280 (1989).
5. Takahata, N. & Nei, M. *Genetics* **124**, 967–978 (1990).
6. Hill, A. V. S. in *Molecular Evolution of the MHC* (eds Klein, J. & Klein, D.) (Springer, Heidelberg, 1991).
7. Conway, D., Greenwood, B. M. & McBride, J. S. *Parasitology* **103**, 1–6 (1991).

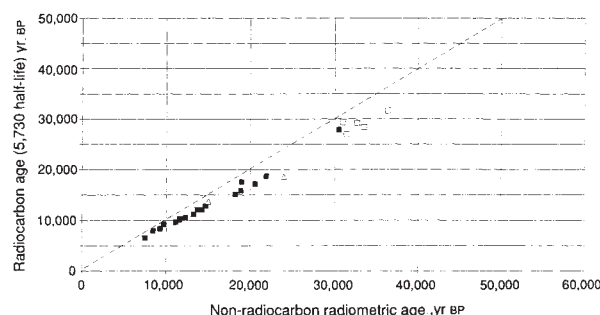
## Language gap

SIR — Foley in News and Views<sup>1</sup> says he finds it "improbable that the early hominids did not have some form of language". This leads to scepticism about the apparent gap between dates for the appearance of anatomically modern *Homo sapiens* and our estimate<sup>2</sup> for language emergence. Foley does not say what hominid "prototype" language might be: there is no easy path from vocal (or any other bodily gestural) communication to language. Language entails perception of the meaningfulness of communicative gestures, whose property of conveying meaning had to have been discovered. Vervets find their own utterances unremarkable<sup>3</sup>, but once hominid ancestors discovered that meaning was what their signs had, these could become themselves objects of perception.

Modern skeletal morphology must appear before the behavioural manifestations

of language in an evolutionary argument. Morphological variation is not 'driven' but, appearing by mutation and genetic recombination, is then selected naturally. Modern anatomy may have been selectively advantageous initially because it permitted more efficient breathing during strenuous activity<sup>4</sup>, carrying with it a capacity for more flexible vocal utterance — not, we stress, the same as language.

In our paper<sup>2</sup>, we pointed to the first



Variation of radiocarbon estimates with age before present estimated by thorium/uranium dating<sup>9,10</sup>, or thermoluminescence dating<sup>5,9,10</sup>. (Black squares, ref. 10; open squares, ref. 9; open triangles, ref. 5.) Estimates of error are not shown because the issue is the trend shown by the central values for independent series of paired estimates.



colonization of Australia more than 40,000 years ago as the earliest indication of language-based skills. Recent results appear to put the event before 53,000 years<sup>5,6</sup>. Theoretical and empirical evidence shows that radiocarbon substantially underestimates ages even beyond tree-ring calibrations<sup>7,8</sup>. Paired radiocarbon and other radiometric ages (see figure) show a consistent trend, predicting that conventional radiocarbon estimates of 40,000 years for first colonization of Australia approach the current claim, by thermoluminescence, of 53,000 years.

Archaeology is the only means of calibrating rates of hominid evolution. There are no independent criteria for estimating the time gap between modern human morphology and modern human behaviour. We need a clear perception of what lies on either side of the gap.

IAIN DAVIDSON

Department of Archaeology and  
Palaeoanthropology,

WILLIAM NOBLE

Department of Psychology,  
University of New England, Armidale,  
New South Wales 2351, Australia

SIR — There is a striking correspondence between 'phylogenetic' trees of human linguistics and those of human genetics<sup>11</sup>. This has been used in a News and Views article<sup>1</sup> to support the view that both trees reflect the historical divergence of human populations. Leaving aside the many general problems associated with the Hennigian reconstruction of phylogeny<sup>12</sup>, there is a particular problem here: the method assumes that there is no interchange between branches of the tree once they have separated, yet human populations that are very different both genetically and linguistically may exchange both genes and language components. Knowing insufficient about language to be sure, I wonder whether such exchange may not be a major reason why the so-called phylogenetic trees are so similar.

But I am quite sure that the graph in Foley's article<sup>1</sup> used to illustrate an aspect of the correspondence between language and genetics presents nothing more than a statistical artefact. It shows a strong correlation between the number of languages in human groups and the between-subgroup genetic diversity within those groups. But only four of the eight groups presented are statistically independent; the others are formed by successive hierarchical grouping of these (and of a ninth group for which no data are shown). Unless two groups do not differ in the variation within them, a supergroup formed by combining them will show more variation than either of them in all characteristics, so the inclusion of supergroups (and supersuper-

groups) on the same graph as groups when looking at the relationship between two sorts of variation is bound to produce an apparent correlation. In this case, if one eliminates the four supergroups from the graph, the apparent correlation is entirely eliminated.

JEREMY J. D. GREENWOOD

British Trust for Ornithology,  
The Nunnery, Nunnery Place,  
Thetford, Norfolk IP24 2PU, UK

FOLEY REPLIES — One of the clearest points made by Noble and Davidson in their paper<sup>2</sup> was that the presence of language should be directly rather than indirectly inferred, hence their stress on the 40,000-year datum. Their arguments relating to Australian colonization seem at variance with this position. Either Noble and Davidson are recoupling behavioural and anatomical evolution, arguing that because we know only anatomically modern humans reached Australia, and that this event was at least 53,000 years ago, therefore language must be older still; or else that communication is essential for such a journey, which is special pleading for humans given the occurrence of colonization events such as playrrhine monkeys in the New World in the Oligocene. Such special pleading is closer to the traditional approaches that they were at pains to resist.

Personally, I am perfectly happy to believe that hominids have been making both remarkable and unremarkable utterances for the past 100,000 years. Definitions of language that are not simply tautologically self-referential to humans remain a problem, but such definitional nuances do not circumvent the problem of what earlier and contemporary (Neanderthal) hominids with very high encephalization quotients were doing with their brains. My suggestion was that distinguishing language as the basis for thought from language as communication might provide a way forward that is consistent with both ethological and archaeological data. Far from bypassing the issue of proto-language, this implies that the structure of what became communicated language evolved as systems of thought that were then grafted on to animal communication systems through the evolution of speech.

Turning to Greenwood's letter, it is perhaps a truism to state that the problem in investigating the biology of human evolution is that there is only one species. This imposes major methodological problems, particularly when we attempt to extend analyses into extrasomatic patterns such as language, and Greenwood rightly draws attention to some of these in my News and Views article. In particular they arise from applying methods developed for inter-

specific analysis to within-species variation. In considering the relationship between linguistic and genetic variation there are no clear-cut clades, even at the terminal twigs of any branching model, so that even a plot of independent data points, as opposed to the hierarchically organized ones that were used, is unlikely to be truly independent. The purpose of the plot was not to read too much into the significance levels but to examine at a broad level the comparative rates of genetic and linguistic divergence. Whether there is a linear relationship in such data is a matter of probability, not inevitability, which affects the interpretation of the statistics but not the validity of the general observation. As I stated in News and Views<sup>1</sup>, there are good historical and geographical reasons why more 'recent' clades should show greater variation. Over time these may be evened out. This effect can be seen by plotting the limited data on genetic distance and language diversity<sup>11,13</sup> at higher (continental) through to lower (regional) levels; the higher the level, the greater the relationship, the lower the greater the scatter. Regrettably, of course, the higher the level the smaller the sample size and hence the lower the significance. There is, though, some evidence in microevolutionary studies<sup>14</sup> for such a positive relationship to exist.

In the end we are left with the observation that there is some relationship between genetic and linguistic divergence. To determine whether this relationship is the product of common ancestry as I implied or subsequent population interaction, as Greenwood favours, requires more formal models that might show how these two interpretations would differ empirically. In the absence of such formal models all we have at the moment are the observations and some *ad hoc* generalizations (my own included).

ROBERT FOLEY

Department of Biological Anthropology,  
University of Cambridge,  
Cambridge CB2 3DZ, UK

1. Foley, R. A. *Nature* **353**, 114–115 (1991).
2. Noble, W. & Davidson, I. *Man J. R. Anthropol. Inst.* **26**, 223–253 (1991).
3. Cheney, D. L. & Seyfarth, R. M. *How Monkeys See the World* (Chicago Univ. Press, 1990).
4. Lieberman, P. *The Biology and Evolution of Language* (Harvard Univ. Press, Cambridge, 1984).
5. Roberts, R. G., Jones, R. C. & Smith, M. A. *Nature* **345**, 153 (1990).
6. Hiscock, P. *Archaeol. Oceania* **25**, 122–4 (1990).
7. Barbetti, M. & Flude, K. *Nature* **279**, 202–205 (1979).
8. Stuiver, M. et al. *Quat. Res.* **35**, 1–24 (1991).
9. Bell, W. T. *Archaeometry* **33**, 43–50 (1991).
10. Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. *Nature* **345**, 405–410 (1990).
11. Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. & Mountain, J. *Proc. natn. Acad. sci. U.S.A.* **85**, 6002–6006 (1988).
12. Benton, J. J. *Trends Ecol. Evol.* **5**, 393–394 (1990).
13. Ruhlen, M. *A Guide to the World's Languages* (Stanford University Press, 1987).
14. Neel, J. V., Layriss, M. & Salzano, F. M. in *Population Structure and Human Variation* (ed. Harrison, G. A.) 109–142 (Cambridge University Press, 1977).

# Imperious innovator

Leonard Krishtalka

**An Agenda for Antiquity: Henry Fairfield Osborn and Vertebrate Paleontology at the American Museum of Natural History, 1890–1935.** By Ronald Rainger. *University of Alabama Press: 1991. Pp. 360. \$37.95.*

FOR almost half a century, Henry Fairfield Osborn (1857–1935) was the archduke of vertebrate palaeontology in North America and the architect of natural science at the American Museum of Natural History in New York. After a college education at Princeton, Osborn's conquest of the museum was swift and durable: curator of palaeontology in 1891, vice-president in 1901 and president from 1908 to 1933. During that period, Osborn tried to create the museum in his own image — few would deny that he was an arrogant, bigoted tyrant. What Osborn created is (arguably) the most prominent and influential natural history institution in North America. Rainger's account of Osborn's rule is the first scholarly history of a fateful period in North American vertebrate palaeontology. From 1890, with the deaths of the two rival palaeontologists E. D. Cope and O. C. Marsh, until the 1930s, the profession slowly abandoned its feudal autocracy for a more egalitarian order. Central to that evolution were Osborn and the museum.

Two events, separated by almost one hundred years, begin to reveal what Rainger, a historian of science at Texas Tech University, calls Osborn's 'agenda for antiquity'. One occurred last month when the museum unveiled its newest dinosaur spectacle: a monstrous skeleton of a *Barosaurus* reared up on its hind limbs, its long neck and peewee head disappearing 50 feet up into the stratosphere of the rotunda. The posture of the barosaur, if not blessed by science, is blessed by the legacy of Osborn's exhibitions: the victory of ideology over the evidence. Here, chemistry and steel is made to do what muscle and bone could not: the living, teetering barosaur, 150 million years ago, did not have the structural benefits of a lightweight fibreglass skeleton or an anchor bolted to its backbone. Similarly, as Rainger shows, Osborn promoted museum attendance and his own beliefs on evolution and human origins with provocative, although scientifically spurious, museum displays.

An old black-and-white print records the second event, a group shot taken at Sheep Creek, Wyoming, a few miles from the museum's extensive dinosaur quarries at Como Bluff. It is August 1899, and Osborn and the museum's field party have come to inspect the

remarkable discovery of two *Diplodocus* skeletons by the new, upstart Carnegie Museum in Pittsburgh. In the Sheep Creek quarry is the turn-of-the-century dream-team of palaeontology: from the American Museum, Osborn, William



Dinosaur-digging dream-team at Sheep Creek in 1899. Osborn is fourth from left.

Diller Matthew, Walter Granger and Richard Swann Lull; from the Carnegie Museum, director William J. Holland, William Reed and Jacob L. Wortman, an embittered former employee of Osborn. Only three luminaries of the time are absent: William K. Gregory, who, not being a field man, is back at the American Museum; and John Bell Hatcher and Barnum Brown, who are in Patagonia leading the third and final Princeton Expedition. The photograph captures the first item on Osborn's agenda: massive, worldwide, expeditions to collect dinosaurs and other beasts for research and exhibition at the American Museum, an endeavour in which he plotted to outdo rival institutions, such as Princeton University and the Carnegie Museum.

How did Osborn complete his agenda? Rainger presents a five-point thesis. First, vertebrate palaeontology was an expensive, peripheral discipline at the turn of the century, supplanted in popularity and prestige by the new experimental sciences of genetics, embryology and biochemistry. Second, to overcome that handicap, Osborn used his force of personality and upper-class social standing to solicit funds for the museum's expeditions, research programmes and exhibits from wealthy, prominent New York families such as the Morgans (J. P. Morgan Jr was

Osborn's cousin), Dodges, Jesups and Fricks. Third, Osborn's success with these patrons reflected their shared, upper-class sociocultural and political objectives, namely, the social and spiritual stewardship of the masses, and the preservation of nature, class and race. Fourth, Osborn tailored the palaeontological evidence to fit these views and values in his voluminous writings and in the enormous museum exhibitions and dioramas. Osborn advocated an orthogenetic, purposeful view of biotic evolution and, as Haeckel did, a racist view of human evolution. Fifth, in vertebrate palaeontology, Osborn had inher-

ited a discipline that operated through a three-tiered social pyramid: at the top, the prominent philanthropists, such as Morgan and Carnegie; in the middle, a socially acceptable, upper-class scientist-manager, such as Osborn or Holland; at the bottom, the low-paid field workers, the bonehunters, such as Hatcher, Wortman and Granger, who more often than not were kept from studying and publishing on the vast collections they had made. Here, Osborn relaxed the class lines drawn by Cope and Marsh. He launched his students, Matthew and Gregory, into careers, but their research areas were by his bidding: biostratigraphy, palaeobiogeography, functional morphology, primate and human evolution. And, despite his conceit, Osborn respected intellectual pluralism, supporting Matthew and Gregory even though they sharply criticized his views on evolution and race.

History involves narrative and cause. Rainger's is a sweeping narrative, from the politics and personae of dinosaur palaeontology to high finance, high society and the eugenicist Galton Society in New York. He dislikes Osborn — there is much to dislike — but tries hard not to show it, giving grudging credit where due. When it comes to cause, Rainger attributes Osborn's agenda and success to upper-class, economic determinants. Implicit in the argument is the notion



that the motives of the wealthy were the worst. Philanthropy fulfilled the social and moral wishes of the upper class. International fossil-hunting expeditions fulfilled America's imperialist ambitions. And the museum's exhibits, guided by Osborn, fulfilled his bigoted notions of evolution and race. The latter is true, but the rest leaves me sceptical.

Historians tend to find causes where they look for them — for Rainger, they are in the middle and upper strata of palaeontology's social pyramid. But in addition to supporting the American Museum, J. P. Morgan, for example, gave generously to the Metropolitan Museum of Art, Harvard University, the Lying-in Hospital of New York and New York trade schools — a roster that implies beneficence rather than the preservation of an elitist status quo. For Andrew Carnegie, a philanthropist of peasant stock, his Pittsburgh museum was "my monument, because here I lived my early life and made my start, and I am to-day in heart a devoted son of dear old smoky Pittsburgh". No elitist aims here either. There are also some anomalies. For example, if Osborn had J. P. Morgan in his hip pocket, why did he lose the Patagonian expeditions (which Morgan funded) to Princeton? Perhaps history, like natural populations, suffers random variations.

No matter. This is an excellent, dis-

passionate history of palaeontology and empire at the American Museum of Natural History. Rainger captures the institution, the era, the players, the stakes, the issues and the darker motives. A final, sad irony of the Osborn saga occurred six years after Rainger's history ends. In 1941, Barnum Brown shipped the type specimen of *Tyrannosaurus rex*, one of Osborn's prized possessions, from the American Museum to the Carnegie Museum where the skeleton still presides over Dinosaur Hall. Brown, Osborn's *alter ego*, had collected the dinosaur in Montana in 1902. The reason for the transfer, says the American Museum, was to keep the skeleton safe in case the *Luftwaffe* bombed New York. But the truth, buried in Barnum Brown's correspondence, is less patriotic. In 1941, with Osborn dead, the American Museum's department of vertebrate palaeontology was near bankruptcy, thanks to a backlash to Osborn's rule. Brown sold the *Tyrannosaurus* to the Carnegie Museum for \$7,000 to establish a modest endowment for the department. The skeleton arrived at the museum three days before the bombing of Pearl Harbor. □

Leonard Krishtalka is at The Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA.

## Duplicating success

Sue Wickner

**DNA Replication, Second Edition.** By Arthur Kornberg and Tania Baker. W. H. Freeman : 1991. Pp. 931. £49.95, \$69.95.

THE first edition of *DNA Replication* and its predecessor, *DNA Synthesis*, both by Arthur Kornberg, made a considerable impression by providing authoritative yet readable summaries of the nucleic-acid enzymology behind the copying of DNA during replication. The new edition is even better. It has been completely revised and updated, and has increased in both scope and depth.

Kornberg is professor emeritus in the department of biochemistry at Stanford University. He won the Nobel prize in 1959 for his discovery of DNA polymerase I and has remained a leader in the rich field of DNA replication. Baker studied for her PhD with Kornberg, currently works at the National Institutes of Health and is affiliated to the department of biology at the Massachusetts Institute of Technology. Together they have created a text that should become a classic in science.

*DNA Replication* is much more than a summary of the field: it is an excellent reference book on DNA and nucleic-acid enzymes, and every molecular biology laboratory should have a copy for easy access to the principles underlying today's DNA technology. When *DNA Synthesis* was published in 1974, DNA replication was investigated almost entirely by scientists interested in understanding the enzymes involved in replication and the process of heredity, and it was those scientists who provided much of the foundation for the DNA technology that has revolutionized molecular biology. Now, almost 20 years later, researchers in all fields of biological science are using these nucleic-acid enzymes.

The book is an excellent resource for graduate teaching and should also be a useful source of background material for undergraduate courses in biochemistry. By putting findings in a historical setting, the authors allow the reader to see how the subject has developed and to understand the pivotal experiments. Topics covered include DNA structure, DNA metabolism, replication enzymes and specific replication mechanisms of bacteria, plasmids, phages, animal viruses and animal cells. In areas that I know best, I was impressed by the accuracy of the information and the thoughtfulness of the discussions. The reference list is



FLY River trumpet-eared bat (*Kerivoula muscina*). There is little information on the habits and life history of this rare New Guinean mammal, which has been recorded on only four occasions. Like other members of the genus, however, it would appear to roost singly or in groups of as many as six individuals, often in tree hollows and foliage. The orange oval-shaped areas on either side of their rostrum are caused by lenses of subcutaneous fat or glands, the function of which is unknown. This picture is taken from *Mammals of New Guinea* by T. Flannery. Published by Robert Hale, £40.

well balanced, neither too long nor too short. The chapters are independent so the reader can jump into the book at any point; at the same time, many unifying themes connect the book's different parts.

The figures are exceptional, surpassing those in the previous edition. They do more than summarize the text; they allow the reader to visualize complicated concepts. For example, drawings efficiently communicate the multiprotein interactions in initiation complexes at origins of replication and in DNA

elongation complexes at replication forks. By comparison to the first edition, the figures alone show how much clearer the mechanisms of DNA replication have become in the past decade.

My only regret is that the expansion of the book has rendered Kornberg's personal vision less apparent, although thankfully his well-known love of enzymes still shines through everywhere. □

*Sue Wickner is in the Laboratory of Molecular Biology, National Institutes of Health, Bethesda, Maryland 20892, USA.*

## Coffee-table cosmology

Joseph Silk

**The Natural History of the Universe.** By Colin A. Ronan. *Doubleday/Macmillan Inc.* 1991. Pp. 212. £20, \$39.95.

**Ancient Light: Our Changing View of the Universe.** By Alan Lightman. *Harvard University Press.* 1991. Pp. 170. \$18.95, £15.25.

**Journey Through the Universe.** By Jay M. Pasachoff. *W. B. Saunders.* 1991. Pp. 391. £25.95, \$38.75.

**Isaac Asimov's Guide to Earth and Space.** By Isaac Asimov. *Random House.* 1991. Pp. 285. \$20.

THE Big Bang theory is our modern mythology, replete with scientific credentials of the highest order. In all environments, from formal banquets and cocktail bars to academic classrooms and funding agencies, it provides an endless source of debate and inspiration. Popular books continue to appear that attempt to communicate the fascination and excitement of the quest for our origins that underlies the science of cosmology.

Ronan's *The Natural History of the Universe* is a glossy coffee-table book that is a worthy addition to science popularization. An enthusiastic amateur astronomer, Ronan has generated a beautifully produced and illustrated sequence of thumbnail sketches encompassing the entire scheme of cosmic evolution. This splendid overview cannot fail to impress, despite occasional flaws in detail. Professional astronomers are woefully derelict in their duty to popularize, yet would not have been quite so vulnerable as Ronan to the latest press releases. Recent discoveries of low-mass stars have failed to provide any indication of the nature of "missing" mass. Nor would any astronomer tolerate "unknown numbers of black holes of unknown mass" as an alternative. One can only become more mystified as to what 'texture' might be after reading Ronan's

book (it's a kind of fossil froth caused by the Big Bang), and it is not true that inflation explains the observed ratio of photons per baryon. But these are quibbles: let the reader judge.

In *Ancient Light*, Lightman demonstrates the reliability of a professional astrophysicist. Inevitably this slender account of the Big Bang theory is more technical than Ronan's book, but is enlivened by portraits of the leading characters. Due credit is gratefully given to the many creators of inflationary cosmology, although Andrei Sakharov is unaccountably omitted as the godfather of baryon synthesis. We learn that the low entropy inferred before inflation at the beginning of the Universe is about as unnatural as finding the outcome of wiring one's favourite pet to a computer terminal to be a Beethoven sonata. But we also know that the quantum gravity theory of the first  $10^{-43}$  seconds of the Universe is so woefully inadequate that any extrapolation of the second law of thermodynamics — that entropy never decreases — becomes highly dubious. This is more a flaw in our fundamental physics than in the Big Bang theory.

For the briefest instant, lasting less than  $10^{-30}$  seconds, the very early Universe underwent a dramatic increase in its rate of expansion. Explaining inflation provides an acid test that few popularizations pass. Lightman talks about a false vacuum that acts like negative gravity, repelling matter. Ronan tells us that the false vacuum forms when supercooling occurs (in fact, it was always present, but only briefly did it come to dominate the pressure). One could circumvent some of the jargon by noting that cooling as the Universe underwent a phase-change temporarily results in a negative pressure as regions form of the new phase: low-energy examples are condensation of steam, freezing of water and formation of crystals. Negative pressure is negative energy density, which is negative mass, which in turn is antigravity: hence repulsion. This is what drives inflation. The mystery, still unsolved, is why the pressure remained negative long enough to generate the enormous degree of inflation that is required, followed by

the relatively smooth reheating of the now supercooled Universe to reproduce the hot Big Bang that is a cherished article of faith among cosmologists.

Those looking for a more academic approach to the story of everything will benefit from Pasachoff's *Journey Through the Universe*. Pasachoff has distilled the cosmic essence from his elementary astronomy textbook. The result is a lavishly illustrated and scientifically honest description of contemporary astronomy and contemporary astronomers. Isaac Asimov's *Guide to Earth and Space* takes an alternative tack. Asimov poses more than one hundred questions that span properties of the Earth to stars and galaxies, culminating in the usual "Where are we going?" and "Where did we come from?" His answers are concise and nontechnical, a boon for the reader in a hurry. Yet when on unfamiliar ground — inflation again — he too pays incoherent tribute to fluctuations in a false vacuum.

None of these attempts at popularization need disgrace a scientist's bookshelf. All are carefully researched and well written. Perhaps the touch of inspiration is lacking, but the intended audience of interested but scientifically semiliterate readers will surely benefit. □

*Joseph Silk is in the Departments of Astronomy and Physics, University of California, Berkeley, California 94720, USA.*

## Methods books

■ There are two new introductions to the handling of numerical data: *A Practical Guide to Data Analysis for Physical Science Students* by Louis Lyons (Cambridge University Press, price £19.95, \$34.95 (hbk), £7.97, \$12.95 (pbk)) and *Basic Statistics for Laboratories: A Primer for Laboratory Workers* by W. D. Kelly, T. A. Ratliff Jr and C. Nenadic (Van Nostrand Reinhold, price \$44.95).

■ *Radioisotopic Methods for Biological and Medical Research* by H. W. Knoche is published by Oxford University Press, price £30.

■ *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules* by D. Lin-Vien, N. B. Colthup, W. G. Fateley and J. G. Grasselli is sure to become a standard reference work for every practicing spectroscopist. Published by Academic, price \$89.95.

■ IRL Press at Oxford University Press have issued two new volumes in their successful Practical Approach Series. *Essential Molecular Biology*, edited by T. A. Brown, covers the procedures needed to identify specific genes and study their structure and cellular expression. *Oligonucleotides and Analogues*, edited by F. Eckstein, will be of interest to academic and industrial researchers. Each is issued in paperback at £22.50, but is also available as a spiral-bound hardback.

■ New in paperback is *A Handbook of Silicate Rock Analysis* by P. J. Potts. Published by Blackie, price is £45. For a review see *Nature* **327**, 198 (1987).



# Unknotting epistemology

John Ziman

**The Philosophy of Science.** Edited by Richard Boyd, Philip Gasper and J. D. Trout. MIT Press: 1991. Pp. 800. £44.95, \$50 (hbk); £22.50, \$25 (pbk).

IT is unlikely that many readers of *Nature* will want to buy an anthology of 41 famous articles on the philosophy of science, even with seven Introductory Essays by the editors. But the wide range of authors includes many creative scholars who are still very active, such as Evelyn Fox Keller and Daniel Dennet, with whose ideas a serious scientific intellectual ought surely to become acquainted. In that spirit, I could recommend this massive volume to the science-faculty librarian, even without reading it.

To say anything more, I ought to plunge into the text — a formidable labour whose products could not possibly be presented in the space available here. The only person with the expertise to review the compilation as a whole would be someone who had spent as much time as the editors in collecting and interpreting just such material. Even an assessment of one of the seven sections, for example that of the five papers representing the philosophy of physics, would require specialized knowledge of a large body of writing, much of which is still highly contentious.

So let me approach the book in another way by asking what an experienced professional natural scientist might learn about his or her calling from dipping into it. I will ignore Part II, which deals with the various philosophies of various great disciplines — Physics, Biology, Psychology and Social Science — on the assumption that my hypothetical browsers have already reflected far more deeply on the fundamentals of their own subjects than could be said in 80 pages or so. Let us concentrate on Part I (24 chapters, 460 pages), which is about epistemology: what do we really know, or what do we think we know, when we think we know something 'scientifically'?

An anthology inevitably contains much that is primarily of historical interest. Philosophers insist that their best work does not date; but is the philosophy of science firmly rooted in perennial problems? Apart from a familiar but misplaced extract from Thomas Kuhn, I found myself flipping over the first 260 pages with scarcely a pause to read more closely. The key words in the chapter headings — "confirmation", "realism", "empiricism", "falsification", "explana-

tion" and so on — said enough. We were still in the post-scholastic era, when it seemed essential to (dis)prove the peculiar (un)reality of scientific knowledge (theories/facts/data/hypotheses) by analysing (deconstructing) the arguments on which it was (supposedly) based. But as David Hume demonstrated more than 200 years ago, this is an exercise in circle-squaring, whatever way you want it to come out. Scientists scent this intuitively, and are repelled by this sort of philosophy.

Suddenly, in chapter 14, written by Arthur Fine in 1984, we find ourselves in a world we can recognize. It turns out that what he calls the "natural ontological attitude" — that is, common sense, everyday, pragmatic realism, with no absolute pretensions — is OK after all, and makes as solid a foundation for an epistemology of science as any we could ever construct. In other words, scientific entities are neither more nor less 'real' in principle than the rest of the 'external world', whose existence we take on trust, whatever the philosophers say about it. Incidentally, we read that Morris Schlick had said something like this in 1932, but it must have seemed much too naive, much too much like 'folk-philosophizing', for the professionals.

Unfortunately, this is where the track to a more satisfactory philosophy of science peters out. Surprisingly, Fine does not mention the German phenomenologists of the early twentieth century (especially Alfred Schultz) who went a long way in analysing the 'natural attitude', even though they did not apply this analysis to scientific thought. The anthology does not even include anything from the post-kuhnian 'sociologists of scientific knowledge', who argue cogently that all science is a social construct, designed to further the interests of particular groups in a struggle for power and influence. This view has eloquent spokesman, such as Harry Collins and Bruno Latour, whose voices should have been heard in this debate. But I do not myself go along with them in their epistemological relativism, not because it is too radical and anarchic, but because they have not got the courage to apply it to all human knowledge, including what we think we know about the everyday world.

Indeed, if my hypothetical browser is really interested in pursuing such matters further, the next step might be out of self-styled philosophy of science into a more general philosophical realm, where there is now a tremendous ferment of new ideas. I do not pretend myself to be seriously informed about these, but they are most lucidly and elegantly summarized by Don Cupitt in *Creation out of Nothing* (SCM Press, 1990) as a foundation for a new theology. In effect, there

is no logical counter to nietzschean nihilism about this or any other world, outside our own selves. But we exist as conscious beings through language, projecting its categories and implicit relationships on whatever it is that we conceive this 'world' to be. Because natural languages are social institutions and differ markedly from culture to culture, the relativism of the sociologists of scientific knowledge is well justified.

Or is it? My own path, strangely enough, would now take me back into all that old stuff about 'confirmation', 'explanation', 'hypothesis' and so on, and I would ask why it seemed relevant to scientific activity. One could then see these as signs of a long-term collective effort to create a universal language — in its ideal form, mathematical — that would transcend the relativist Babel. The demand for 'objectivity', for example, can no longer be given absolute significance, but would still have meaning as a striving for consensus among diverse observers. Again, factual or formal contradictions are scientifically abhorrent, not because 'the world' is ultimately 'logical', but because such contradictions introduce ambiguities in what scientists try to tell each other about their experiences.

Needless to say, such an account of science would still remain open to the nihilist critique. As the mathematicians and intelligence artificers have discovered in recent years, the notion of a universal, unambiguous consensual language is a chimera, even for saying  $2 + 2 = 4$ , let alone for the construction of an all-wise computer. But one can see how it is that scientific knowledge, although not uniquely privileged as 'the truth', is different from other bodies of organized knowledge, such as aesthetics, law or religion, where intellectual consensus is not the primary objective.

As this anthology demonstrates, established notions of the philosophy of science almost completely ignore its social dimensions. The philosophers see research and discovery primarily as the activity of individuals, instead of seeing them as the collective work of a community. Their nets are all knots, connected by very tenuous strings: they forget that a net is only as strong as its links, and that the enclosed public space is where the action is. Working scientists understand this instinctively. In spite of their rugged individualism, they take pride in the collectivity of their enterprise, which actually makes it seem part of the nature that they delight in exploring. □

John Ziman is emeritus professor of physics at Bristol University, H. H. Wills Physics Laboratory, Royal Fort, Tyndall Avenue, Bristol BS8 1TL, UK.

# Molecular dissection of the secretory pathway

James E. Rothman & Lelio Orci

A combination of biochemistry in animal cell-free systems and genetics in yeast is revealing the molecular machinery of the secretory pathway of eukaryotes. Transporting vesicles have a simple coat structure and employ a general mechanism for fusion that is conserved in evolution.

THE secretory pathway of eukaryotic cells underlies the targeting of newly synthesized proteins, the generation and maintenance of subcellular compartments and organelles, and proper cell growth and division. Insight into this complex maze of vesicular transport pathways has come in several waves, beginning with the pioneering work of the 1960s (ref. 1) in which the fundamental relationship of the endoplasmic reticulum (ER), Golgi and secretory storage vesicles in regulated secretion was deduced, and the need for vesicular carriers to ferry cargo between topologically separate compartments was first recognized. These now central examples of cell biology were extended in the 1970s to the formation of plasma membranes, lysosomes and other organelles, and continue to be refined by ever more sophisticated techniques of immunocytochemistry. In the 1980s, elucidation of the molecular machinery of vesicular transport began with the reconstitution of intracellular transport in cell-free systems from animal cells<sup>2</sup> and the isolation of secretory mutants in yeast<sup>3</sup>.

Genetics and biochemistry are teaching us that many of the components of the secretory pathway are universal: the same machinery operates in yeast as in animals. A key enzyme for fusion of a vesicle with the Golgi is also needed for fusion of endocytic vesicles. The coats that pinch off endocytic and Golgi vesicles have underlying similarities not evident from their morphology. And we learn that members of gene families (like the small GTP-binding proteins ('smgs')) are modular units that perform the same tasks in different places, combining a general mechanism with compartmental specificity. These and other themes, along with many important details, continue to emerge. Here we will focus on the transport pathways that operate in the ER and Golgi, and only touch on the related subjects of secretory granule formation, regulated exocytosis, lysosome biogenesis, and sorting to apical and basolateral domains in polarized cells.

## Compartmental and information organization

The passage of a freshly translocated protein from the rough ER to the cell surface involves sequential transit through a series of topologically distinct (but physically associated) compartments collectively referred to as the Golgi complex (Fig. 1). This intracellular transport is mediated by a series of carrier vesicles that bud from each compartment in the series and then fuse with the next, giving rise to vectorial movement.

The Golgi is organized into three functionally distinct regions (Fig. 1), the *cis*-Golgi network (CGN), the Golgi stack (containing several distinct subcompartments), and the *trans*-Golgi network (TGN). The *cis*- and *trans*-Golgi networks are the entry and exit faces of the stack, respectively, and are mainly sorting and distribution centres. The CGN serves, at least in part, as a filter to remove certain escaped ER proteins (see below). The TGN is where proteins with differing final destinations, such as lysosomes, secretory storage vesicle and plasma membrane domains diverge<sup>4</sup>. The stack of cisternae between them functions as a series of distinct processing stations in which, for example, the saccharide side chains of transported glycoproteins are successively matured. There is some debate as to the exact

number of topologically separate compartments in the stack, but most evidence suggests there are three, termed *cis*, medial and *trans* in the order of their encounter. Transit of proteins between compartments seems to be unidirectional, mediated by transport vesicles<sup>5</sup>.

Supporting this model of the Golgi are many studies in which individual processing enzymes, such as glycosyltransferases and less well defined antigens, have been mapped by electron microscope immunocytochemistry and subcellular fractionation, along with transported proteins<sup>6-8</sup>. Temperature blocks cause accumulation of transported proteins at particular stations of the Golgi, thereby revealing them as distinct. For example, lowering the temperature to 20 °C greatly slows exit from the TGN, and a variety of transported proteins pile up there<sup>4</sup>. Lowering the temperature to about 15 °C causes an earlier block, in which transported proteins accumulate in an expanded CGN compartment<sup>9,10</sup>.

The CGN (defined in refs 11-13) also has a number of other names, including the 'salvage compartment' and the 'intermediate compartment' reflecting a measure of controversy as to how to classify it with respect to the Golgi stack. This also reflects the differing emphasis of different investigators on different aspects of this compartment's function. If the Golgi is defined in the most classical morphological sense as a stack of flattened cisternae, then it is not always clear whether the CGN is a part of the Golgi because the morphology of the CGN can vary from tubular and apparently separate from the stack at one extreme, to flat and cisternal at the other. Though historically correct, such a morphological definition of the Golgi may be so narrow as to miss the point. Rather, if we define the Golgi more conceptually, as the set of topologically separate compartments after the ER but before the final destinations (whatever their

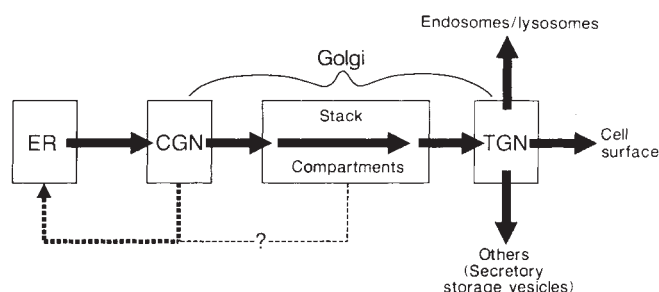


FIG. 1 Flow diagram of the secretory pathway. Thick solid arrows indicate unidirectional forward (anterograde) movement mediated by transport vesicles between successive, topologically separate compartments of the secretory pathway, responsible for cellular membrane growth and biosynthesis. The thick, dashed arrow indicates unidirectional and vesicular retrograde transport from the first compartment Golgi to the ER, selective for ER resident proteins and crucial for their retention. The thin dashed line with the question mark indicates the possibility, for which there is no direct evidence, that retrograde transport can also occur from later Golgi compartments (see also Fig. 5 in ref. 31). The exact number of compartments in the stack is unknown, but is probably three, termed *cis*, medial and *trans*.



shape happens to be under one or another physiological condition), then the CGN stands firmly at the port of entry into the Golgi.

What determines whether a membrane or luminal (soluble) protein in the ER or a Golgi compartment will remain there as a permanent resident, or instead be transported forward ('anterograde') towards the TGN and the cell surface? Compelling evidence exists that anterograde movement is by default, whereas retention requires special signals<sup>14,15</sup>. Proteins unrestrained from free diffusion will enter the walls or lumen of these vesicles according to their prevailing concentration in the appropriate phase, and be swept along with the 'bulk flow' to the cell surface created by the vectorial movement of carrier vesicles in that direction. Thus, cytoplasmic proteins introduced into the ER lumen are efficiently secreted, provided that they can fold and thus bypass the ER's quality control system<sup>16</sup>. Even a simple glycopeptide introduced biosynthetically into the ER is secreted as fast as the fastest protein<sup>17</sup>. When proteins are secreted more slowly than the bulk flow rate, this usually reflects a delay in leaving the ER, probably because of the time required for folding<sup>18</sup>. Proteins targeted to lysosomes, secretory storage vesicles and other locations have sorting signals that divert them from the bulk flow at the TGN (ref. 4).

Several specific and transplantable 'residence' signals have been identified in certain permanent resident proteins of the ER and Golgi. The sequence KDEL (single-letter amino-acid code) at the carboxy terminus of a number of ER luminal proteins<sup>19</sup> indicates residence in the ER. The sequence KKXX (where X is almost any amino acid) at the cytoplasmically exposed C terminus of certain ER membrane proteins also specifies ER residence<sup>20,28</sup>. A hydrophobic spanning segment appears to specify residence in the Golgi stack<sup>21</sup>.

'Residence' signalled by the KDEL sequence is achieved mainly by retrieval of proteins bearing this tag from the CGN ('salvage compartment') if and when they should slip out of the ER along with the bulk flow. The KKXX signal also seems to specify residence by retrieval from CGN and possibly later Golgi compartments (P. Peterson, personal communication). Evidence for a salvage mechanism came from an ingenious experiment using yeast<sup>22</sup>, in which the sequence HDEL (the *Saccharomyces cerevisiae* equivalent of mammalian KDEL) was attached to the C terminus of a protein normally secreted from yeast. The HDEL protein is retained in the ER, yet is also modified by the addition of certain mannose residues characteristic of the early regions of the yeast Golgi. This glycosylation does not occur when ER to Golgi transport is blocked by mutation, confirming that exit from the ER is required to allow modification in the Golgi. KDEL/HDEL- or KKXX-triggered retrieval from the Golgi is by definition selective, and so can occur without causing significant nonspecific reflux from Golgi to ER, and thus without compromising the efficiency of vectorial anterograde bulk flow.

A mutant of yeast defective in retaining HDEL-terminating proteins (erd2) has been identified, and is in the gene encoding specificity for the HDEL salvaging receptor. Thus, expression of the *Kluyveromyces lactis* *erd2* gene in *S. cerevisiae* changed the specificity of the retention mechanism to that of *K. lactis*<sup>23</sup>. The *erd2* gene is predicted to encode a membrane protein of relative molecular mass 26,000 ( $M_r$  26K) that may span the bilayer seven times<sup>24</sup>. A mammalian homologue of *erd2* has been cloned using the polymerase chain reaction (PCR), and encodes a 25K membrane protein that is 50% identical to the yeast receptor and is localized to the Golgi complex as judged by immunofluorescence<sup>24</sup>. Another approach, using anti-idiotypic antibodies, has identified a second putative KDEL receptor from animal cells, a 72K transmembrane protein that is able to bind the KDEL-containing peptides<sup>26</sup>, and which appears to be principally localized to the CGN. Thus, mammalian cells may prove to have two KDEL receptor systems for retrieval.

It would seem rather inefficient to ensure the permanent

TABLE 1 Proteins and genes implicated in budding of coated transport vesicles

| Animal cells         | Yeast cells                                     |
|----------------------|-------------------------------------------------|
| Coatomer             | Budding genes, products and associated proteins |
| $\alpha$ -COP (170K) | <i>sec16</i> gene                               |
| $\beta$ -COP (110K)  | p105 (bound to sec23p)                          |
| $\gamma$ -COP (98K)  | sec23p (85K)                                    |
| $\delta$ -COP (61K)  | <i>sec21</i> gene                               |
| p36 (36K)            | sec13p (34K)                                    |
| p20 (20K)            | sec12p (70K)                                    |
| ARF (21K)            | ARF (21K)                                       |
|                      | SAR1p (21K)                                     |

Other genes and suppressors are known, but have not yet been classified with respect to budding or fusion phenotypes.

residence of ER proteins merely by retrieving them after they have already left the ER. It seems more likely that there will also prove to be other mechanisms for retaining the ER proteins, by actively withholding them from the bulk flow in the first place. Such a 'first line of defence' would greatly reduce the load on the retrograde retrieval pathway, which is in fact easily saturated by overexpression<sup>22,24</sup>. Indeed, a resident protein (BiP), normally terminating in KDEL, is secreted far more slowly than the bulk flow when their KDEL signals are removed<sup>19</sup>, strongly suggesting the existence of an earlier-acting retention mechanism not involving retrieval.

What are the vesicular carriers of the retrograde traffic of KDEL-terminating proteins that have escaped the ER (and possibly other kinds of resident ER proteins)? Studies with the drug brefeldin A (BFA) raise the possibility that retrograde transport is mediated by vesicles in the shape of long (microns), thin (50–100 nm) tubules that bud from the Golgi and move along microtubules until they fuse with the ER<sup>27</sup>. But there is no direct evidence that these membrane tubules, which are involved in the massive reorganization of the Golgi that occurs in the presence of BFA, are responsible for the selective KDEL-type retrograde transport that occurs in the absence of BFA.

How general is retrieval as a residence mechanism for ER proteins? Are all membrane as well as luminal residents of the ER potentially retrievable? How far into the Golgi system of compartments do the ER proteins penetrate? Can retrieval potentially occur from any Golgi compartment? So far, only a few integral membrane proteins of the ER have been shown to escape and then be retrieved<sup>12,28</sup>. Glycoprotein constructs with KDEL termini that are potential recipients of complex glycans added in the medial region of the Golgi stack show no evidence of these modifications<sup>22,29</sup>, nor do most ER membrane proteins analysed as a group<sup>30</sup>. Thus, retrieval under physiological conditions must be very efficient, almost always occurring before exit from the CGN into the remainder of the Golgi. And yet the apparent presence of the mammalian *erd2* KDEL receptor<sup>25</sup> throughout the Golgi (at the level of resolution of light microscopy) and not just the CGN, suggests that the capacity to retrieve may well exist in multiple Golgi compartments. If so, successive Golgi cisternae would serve as a series of potential sites for recapture, so that what few ER proteins may slip past the first Golgi compartment (the CGN) would then be removed from the bulk in the next compartment, and so on (also making post-CGN retrieval steps very difficult to detect).

Taking another kind of retrograde transport, back into the 'dark ages' of the Golgi, this is essentially what was envisioned in 1981 in the 'distillation hypothesis'<sup>31</sup>, which focused attention on the fact that the membrane proteins being exported from the ER (such as precursors to the plasma membrane) constitute only a trace fraction (perhaps only 1/10,000) of the ER resident membrane protein. An extraordinarily efficient sorting of ER

from exported proteins would thus be required to prevent leakage of ER proteins into the Golgi, and the hypothesis was that this degree of sorting could not be achieved in a single step of vesicle budding, necessitating a stochastic process of selective retrieval of ER proteins from the Golgi to improve fidelity. The remarkable observations concerning the KDEL retrieval system, and the effects of BFA, shed new light on this old hypothesis; it will be important to learn just how far into the Golgi the capacity for retrograde transport actually extends.

### Machinery for Golgi vesicular transport

The central mechanistic questions posed by the cell biology just discussed concern how transport vesicles bud and fuse. These questions are the same whether the direction of movement is anterograde or retrograde, and independent of the exact number of transfers needed to cross the Golgi. What force would drive a vesicle to bud off from a membrane? On what basis does each vesicle choose its target for fusion to achieve the specificity inherent in a strictly ordered series of unidirectional transfers? What accounts for the rapid fusion of the transport vesicle with its target membrane, whereas spontaneous fusion between lipid bilayers is extraordinarily slow under physiological conditions? The resolution of these questions in the language of protein-protein interactions constitutes a fundamental understanding of the secretory pathway. The identification and study of the proteins responsible for budding and fusion, on the basis of biochemical approaches using cell-free reconstituted transport systems from animal cells and/or genetic approaches with the yeast *S. cerevisiae*, are rapidly leading to concordant answers.

The molecular nature of the transport vesicles and information concerning the mechanisms of budding and fusion has been most clearly delineated for the process of intercisternal transport in the Golgi stack, and so will be reviewed first, although studies are proceeding rapidly on many other fronts<sup>32,33</sup>. When isolated Golgi stacks are incubated with cytosol (100,000g supernatant) and ATP, numerous ~75-nm diameter vesicles form from every cisterna<sup>34</sup>. These are likely to be transport vesicles, as they contain a transported protein (the vesicular stomatitis virus (VSV)-encoded G protein) present when Golgi fractions are prepared from VSV-infected cells<sup>35</sup>. The concentration of G protein in the 75-nm vesicles is indistinguishable from that of the parental cisternae, and vesicle budding is unlinked to the presence of cargo molecules<sup>36</sup>, both hallmarks expected of carriers of the anterograde bulk flow. Completion of a round of transport of VSV G protein between cisternae can be monitored by transport-coupled glycosylation<sup>36</sup>, that is the maturation of the glycoprotein's saccharide chains after transfer to the next compartment (routinely using *N*-acetylglucosamine as the sugar tag, added in the medial<sup>37</sup> compartment of the Golgi stack). Transport-coupled glycosylation, like the production of 75-nm VSV G protein-containing vesicles, requires both cytosol and ATP. Much evidence implies that transport in the *in vitro* system involves vesicle movements between Golgi stacks that themselves remain intact, and maintains the compartmental specificity that exists *in vivo*<sup>38-40</sup>.

The 75-nm vesicles come in two varieties: those coated on their outer (cytoplasmic) surface with an electron-dense material (not containing clathrin), and those uncoated<sup>35</sup>. The coated vesicles are the precursors of the uncoated vesicles. Transport-coupled glycosylation is blocked by nonhydrolysable analogues of GTP, such as GTP- $\gamma$ -S (ref. 41) and also by pretreatment of the Golgi membranes with the sulphhydryl reagent *N*-ethylmaleimide (NEM) under mild conditions<sup>42</sup>. GTP- $\gamma$ -S blockade leads to the accumulation of buds and coated vesicles<sup>41</sup>, whereas NEM blockade leads to the accumu-

lation of the uncoated variety of 75-nm vesicles<sup>43</sup>. That coated vesicles are precursors of uncoated vesicles follows because biochemical experiments showed that the GTP- $\gamma$ -S block precedes the NEM block to transport, and combining both treatments leads to the exclusive accumulation of coated vesicles<sup>44</sup>.

Altogether, this suggests a pathway (Fig. 2) in which coat proteins assemble on a Golgi cisterna to form coated vesicles which then transfer to the next cisternal compartment, lose their coats (in a reaction that probably requires or is linked to GTP hydrolysis) and then fuse (in a reaction that requires an NEM-sensitive protein). The predicted NEM-sensitive fusion protein (NSF) was purified<sup>45</sup> from cytoplasmic extracts by using its ability to restore transport (for example fusion) to NEM-blocked Golgi membranes as a guide, yielding a homotetramer of 76K subunits that requires ATP for stability. Both budding<sup>46</sup> and fusion<sup>47</sup> require ATP, and each also requires fatty acyl-coenzyme A, for purposes that are still unclear.

Strong correlative evidence links these Golgi (nonclathrin) coated vesicles to the process of transport-coupled glycosylation, which is the biochemical assay used to purify cytoplasmic transport factors. (1) Coated vesicles form under the same conditions as biochemically defined transport<sup>35</sup>. (2) Coated vesicles containing cargo (VSV G protein) transfer from donor to acceptor stacks and are trapped at this stage by the GTP- $\gamma$ -S block, showing that they pinch off and move between cisternae<sup>44</sup>. When this block is reversed (in the case of aluminium fluoride) the accumulated vesicles disappear and VSV G protein is then glycosylated. (3) Transport-coupled glycosylation, like coated vesicle budding, requires fatty acyl-coenzyme A (ref. 46). (4) The accumulation of uncoated vesicles from NEM-blocked Golgi is reversed when pure NSF is added back, and transport-coupled glycosylation resumes<sup>43</sup>. (5) Addition of aluminium fluoride or NEM to whole cells causes the massive build-up of coated or uncoated 75-nm vesicles, respectively, at the Golgi complex *in vivo*<sup>44</sup>. This extensive correlation does not quite prove that the coated vesicles are the intermediates carrying cargo across the stack *in vivo*. Proof awaits the demonstration that the isolated vesicles are competent carriers (difficult because they are isolated after irreversible blockade with GTP- $\gamma$ -S (ref. 41)), that their coat proteins are required for transport *in vitro*, and the identification of a secretion phenotype with mutated

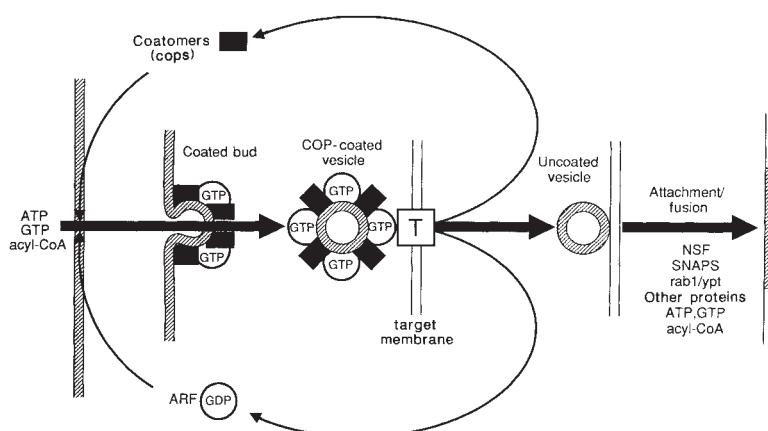


FIG. 2 Working model for vesicular transport between Golgi compartments. Shaded membrane at left represents a portion of a Golgi cisternal membrane giving rise to a nonclathrin COP-coated vesicle which attaches to its target membrane (unshaded), the envelope of the next Golgi compartment in the secretory pathway. Specific targeting of the coated vesicle is proposed to be initiated by an unknown component (T) in the recipient membranes. The vesicle is then uncoated and the general NSF-dependent fusion pathway is triggered, together with attachment of the now uncoated vesicle. Components of the coated vesicle include, but may not be limited to, coatomers (a complex of COPs, filled squares) and ADP-ribosylation factor (ARF), which can exist in both GDP- and GTP-bound forms (circles). The proposed cycle for coatomers and ARF has not been directly demonstrated.



TABLE 2 A partial list of smgs likely to be involved in vesicular transport

|            | Small GTP-binding protein | Location(s) when membrane bound                     | Function(s)                            |
|------------|---------------------------|-----------------------------------------------------|----------------------------------------|
| Rab family | rab1Ap/rab1Bp/ypt         | ER and Golgi <sup>16</sup>                          | Attachment/fusion <sup>80,81,102</sup> |
|            | rab2p                     | CGN <sup>103</sup>                                  | Unknown                                |
|            | rab3Ap                    | Neurosecretory vesicles <sup>104,105</sup>          | Unknown                                |
|            | rab4p                     | Early endosomes, plasma membrane <sup>106</sup>     | Unknown                                |
|            | rab5p                     | Early endosomes, plasma membrane <sup>103</sup>     | Attachment/fusion <sup>107</sup>       |
|            | rab6p                     | Golgi stack <sup>108</sup>                          | Unknown                                |
|            | rab7p                     | Late endosomes <sup>103</sup>                       | Unknown                                |
|            | sec4p                     | Post-Golgi vesicles, plasma membrane <sup>109</sup> | Attachment/fusion <sup>62</sup>        |
| ARF family | ARF1p                     | Golgi stack <sup>54,56</sup>                        | Budding <sup>61</sup>                  |
|            | SAR1p                     | ER <sup>63</sup>                                    | Budding <sup>63</sup>                  |

Numerous additional rab<sup>110,111</sup> and ARF<sup>65</sup> genes have been cloned and sequenced but the proteins have not yet been localized. In most cases, there is a substantial pool of soluble smg protein in addition to the membrane-bound forms.

coated vesicle genes. How strong is the evidence? For comparison, the case of Golgi coated vesicles as carriers seems at least as persuasive as for the established view that cell-surface receptors are endocytosed in clathrin-coated pits<sup>48</sup>.

The logic behind the Golgi coated-vesicle transport pathway, like that of clathrin-coated vesicles, seems clear. The coat probably provides the mechanochemical mechanism to shape the parental membrane into a vesicle. In clathrin-coated vesicles, the coat selects the content<sup>49</sup>. The coat needs to be removed before fusion to allow the vesicle and target bilayers to come into contact. With clathrin-coated vesicles, this seems to occur before docking at the target membrane<sup>50</sup> and is an ATP-dependent reaction catalysed by a constitutively expressed hsp70 member of the heat-shock family of proteins, uncoating ATPase<sup>51</sup>. Golgi-derived coats seem to be removed mainly after contact with the target membrane, and uncoating seems somehow linked to GTP hydrolysis<sup>41,44</sup> though an additional role for uncoating ATPase has not been ruled out. The coat has also been proposed to play an important part in preventing nonspecific fusion<sup>52</sup>.

The Golgi-derived coated vesicles have been purified from salt extracts of Golgi fractions incubated *in vitro* to accumulate the coated vesicle with GTP- $\gamma$ -S (refs 53, 54). They contain a number of major and stoichiometric components that are externally disposed and together comprise the coat (Table 1). These coat proteins ('COPs') include  $\alpha$ -COP (170K),  $\beta$ -COP (110K),  $\gamma$ -COP (98K),  $\delta$ -COP (61K) and a small GTP-binding protein called<sup>55</sup> ADP-ribosylation factor (ARF, 21K), because of its discovery as a cofactor in the cholera toxin-catalysed ADP-ribosylation of stimulatory G protein  $\alpha$ -subunits<sup>56</sup>. ARF was first suggested to be involved in secretion because of secretion mutant phenotypes in yeast and because of its localization (when bound to membranes) to the Golgi stack in animal cells<sup>57</sup>. Immunocytochemistry has localized  $\beta$ -COP and ARF to the coated but not the uncoated 75-nm Golgi-derived vesicles<sup>54,55,58</sup>. The coat is virtually absent from Golgi membranes before *in vitro* transport begins, as little if any  $\beta$ -COP or ARF is present on membranes, with the bulk in the cytosol<sup>35,55,59</sup>. Thus coated vesicles must be assembled from a dispersed pool of cytosolic subunits at the Golgi membrane surface.

Purification<sup>60</sup> of native, cytosolic  $\beta$ -COP revealed that it exists in a complex of 650–700K with other coat proteins. This complex was termed 'coatomer' (short for coat protomer) to connote its likely role as precursor to the Golgi non-clathrin coat. As yet, for technical reasons, there is no direct proof that either ARF or coatomer is required for vesicle budding. Consistent with this expectation, however, are strong pharmacological arguments: BFA (but not a stereoisomer) blocks binding of the coatomer to Golgi membranes and simultaneously blocks coated-vesicle budding<sup>60</sup>. Also, peptides blocking ARF activity block coated-vesicle budding and transport-coupled glycosylation<sup>61</sup>. In addition to  $\alpha$ -,  $\beta$ -,  $\gamma$ - and  $\delta$ -COPs, the coatomer contains two additional subunits of 20K and 36K that are also likely to be

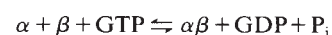
in the coated vesicles<sup>60</sup>. The p20 subunit, despite the similarity in  $M_r$  to ARF, is not ARF; ARF exists separately from coatomer in the cytosol.

An outline of the budding reaction has emerged (Fig. 2), and after the identification of the coat subunits, rapid progress in elucidating details can now be expected with the ability to isolate the coated vesicle product and a cell-free system to generate the vesicles in hand. Coat assembly and budding certainly do not occur by self-assembly, as ATP is required<sup>35</sup>. A role for small (*ras*-related) GTP-binding proteins in secretion was first discovered in pioneering work on the *sec4* gene in yeast<sup>62</sup>. At least one GTP-binding protein now seems to have a crucial role in Golgi coat formation and it is likely that others will be involved<sup>63</sup>.

GTP is required for the coat protein ARF to bind to Golgi membranes<sup>55</sup>. ARF is myristylated at its amino terminus<sup>64</sup>, and myristylation is needed for ARF function in yeast<sup>61</sup>. Myristylated ARF, with GTP bound, inserts spontaneously and non-specifically into pure lipid bilayers; nonmyristylated ARF, or myristylated ARF with GDP bound, does not<sup>65</sup>. This raises the possibility<sup>55</sup> that after a nucleotide exchange or activation at the Golgi surface, ARF is inserted into the Golgi membrane bilayer through the myristic acid, and will remain so until the GTP is hydrolysed, at which time ARF will be released back into the cytoplasm. It is thus tempting to envision, as in our working hypothesis (Fig. 2), a GTP hydrolysis-driven cycle in which ARF binds to the parental membrane as it binds GTP, is incorporated into the coat (perhaps as a major 'integral' membrane protein anchoring the coat) and hydrolyses the bound GTP on encounter with the target membrane, thus destabilizing the coat and recycling ARF and the coatomer to the cytosol. As there is an ever-growing family of ARF-related genes<sup>65</sup>, it is even conceivable that ARF family members carry vesicle targeting information, so that each subtype of COP-coated vesicle might have a different ARF protein to specify its target site for attachment. A specific attachment would then trigger uncoating, and initiate the general NSF-dependent fusion pathway (see below).

The rab family of smg proteins (which includes the *sec4* protein of yeast) are localized to distinct compartments (Table 2), apparently using signals at their extreme C-terminal ends to specify complex patterns of isoprenylation and palmitylation and membrane attachment<sup>66,67</sup>. Their differential localization has given rise to the idea that rab proteins, as a class, bear targeting information specifying where a vesicle should dock<sup>68</sup>. (ARF is not a member of the rab family).

In no case is the exact reaction involving any of the many smg proteins (Table 2) known. But, based on the guiding example of protein synthesis elongation factor Tu, it seems likely<sup>68</sup> that each smg protein catalyses a binding reaction of the kind



where  $\alpha$  and  $\beta$  are two particular macromolecules with affinity

for the given smg. In the case of Tu,  $\alpha$  is aminoacyl-transfer RNA and  $\beta$  is the ribosome. The important point is that in the most general view (Fig. 3), smgs are merely catalysts of pairwise binding equilibria that are driven to completion by GTP hydrolysis (which also regenerates the smg to catalyse another round). Perhaps the large number of smgs associated with secretion simply reflects the large number of protein-protein ligations that need to be made in an orderly and sequential fashion. A few of these may indeed be vesicle-targeting reactions, but it is probably incorrect to assume that every compartmentally localized smg is used in vesicle targeting, as almost every step in a vectorial process such as intracellular transport will have to be spatially organized.

Whereas the smgs may be regarded as some of the 'nuts and bolts' of constitutively operating secretory pathways, recent findings<sup>69,70</sup> strongly suggest that the large, trimeric G proteins<sup>68</sup> may have an important regulatory role, tuning the rate of bulk flow. This may provide a physiologically important means to integrate the rate of vesicle formation and flow with the rate of protein synthesis and the need for cell growth.

Despite the obvious differences between clathrin- and COP-coated vesicles, there may well be underlying similarities. The structure of the clathrin coat is relatively well understood, although little is known of the mechanism of budding from membranes<sup>71</sup>. An outer lattice is composed of clathrin (180K) and two related but distinct light chains (30–40K, depending on the source). The clathrin cage is linked to the enclosed membrane by adaptor particles which have four subunits: two distinct but related 100–115K 'adaptin' chains, one of 47 or 50K, and another of 16K or 19K. The TGN and the plasma membrane have their own distinct set of clathrin adaptors<sup>72</sup>, probably related to the different cargo they select. There is a striking correspondence between the  $M_r$  of the coatomer subunits of COP-coated vesicles and those of clathrin coats<sup>53</sup>. More particularly, the sequence of  $\beta$ -COP is homologous to that of members of the 100–115K clathrin adaptin family<sup>58</sup>. Sequences of other COP proteins are not yet known.

In the simplest view, clathrin and COP-coated vesicles are related structures that form and fuse in related ways. If so, this offers a conceptual simplification that unites the seemingly dis-

parate fields of endocytosis and secretion. Yet, there are fundamental differences in the specificities of the two types of carriers. Clathrin-coated vesicles are selective carriers, executing sorting decisions in which molecules to be removed from the plasma membrane or TGN are selectively concentrated by the coat, whereas resident molecules are excluded. By contrast, COP-coated vesicles seem to be carriers of anterograde bulk flow that do not execute sorting decisions, but rather carry with them whatever proteins diffuse into their walls or lumen by default during budding.

### Fusion of vesicles with cisternae

The coated vesicle is uncoated as or after it docks (Fig. 1), at which point the fusion protein NSF is first obligatory<sup>43</sup>. NSF is a water-soluble protein, lacking conspicuous hydrophobic regions that might trigger bilayer fusion<sup>73</sup>. NSF has a three-domain structure, as predicted from its sequence and confirmed by partial proteolysis mapping<sup>74</sup>. The C-terminal two-thirds of the molecule consists of two direct repeats of homologous ATP domains, each of which possesses a canonical ATP-binding-site motif. Mutating either ATP site eliminates the fusion activity of NSF and reduces ATPase activity. This suggests that ATP hydrolysis by NSF is required for fusion.

NSF requires additional proteins to bind to Golgi membranes and trigger fusion (Fig. 4). Three soluble NSF attachment proteins ('SNAPs') have been purified, termed  $\alpha$ -SNAP (35K),  $\beta$ -SNAP (36K), and  $\gamma$ -SNAP (39K)<sup>75,76</sup>. All three are water-soluble and monomeric proteins, and yet are also hydrophobic in that they bind to phenyl-Sepharose. The most hydrophobic,  $\gamma$ -SNAP, requires 50% ethylene glycol for elution<sup>77</sup>. The apparent amphipathic nature of the SNAPs is noteworthy in light of their proposed role in membrane fusion.  $\alpha$ -SNAP and  $\beta$ -SNAP are structurally very similar molecules, but differ drastically from  $\gamma$ -SNAP in sequence (M. Brunner, D. Clary and J.E.R., unpublished results). An attractive but speculative possibility is that either  $\alpha$ -SNAP or  $\beta$ -SNAP bind to one of the ATP domains of NSF, whereas  $\gamma$ -SNAP binds to the other ATP domain, offering two points of attachment that could be modulated by ATP hydrolysis. SNAP proteins (alone or in combination) do not bind NSF in solution to form stable complexes, and do not stimulate the ATPase activity of NSF<sup>78</sup>. But SNAPs do bind to an integral membrane receptor present in the Golgi membranes, and when bound to membranes are then able to bind NSF<sup>33,75,76</sup>. In Golgi membranes  $\alpha$ - and  $\beta$ -SNAP bind competitively to the same site, whereas  $\gamma$ -SNAP binds independently<sup>78</sup>. These binding reactions are unaffected by ATP. The  $\alpha/\beta$ -SNAP membrane-protein receptor is either the same protein as the  $\gamma$ -SNAP receptor or is noncovalently associated with it, as shown by coprecipitation of bound  $\alpha$ -SNAP and  $\gamma$ -SNAP<sup>78</sup>.

NSF,  $\alpha$ -SNAP,  $\gamma$ -SNAP and the receptor assemble together to form a 20S particle that seems likely to be the core of the active species in membrane fusion<sup>79</sup>. Notably, the 20S particle disassembles in a reaction requiring ATP hydrolysis<sup>79</sup>. Thus ATP hydrolysis by NSF, needed to fuse membranes, at the same time dissociates NSF from membranes. Although the story is far from complete, what clearly emerges is a simple principle by which intracellular membrane fusion is probably regulated, by exploiting controlled assembly and disassembly of a fusogenic particle.

The nature of the problem inherent in controlling intracellular membrane fusion is analogous to that of single photon detection by the visual system. Although rhodopsin must be able to detect a single encounter with a photon with high efficiency, this receptor must not fire in the dark. Similarly, a single receptor-dependent encounter of a target membrane with a transport vesicle must initiate fusion but, equally important, fusion can never be triggered without a targeting event. Suppressing such 'dark reactions' must be crucial; imagine the consequence of even one fusion event between a lysosome and the nucleus in

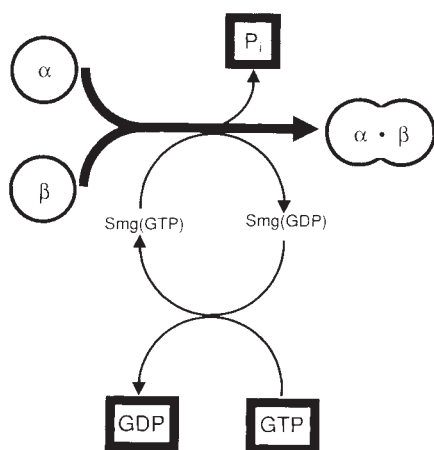


FIG. 3 The general reaction believed to be catalysed by smgs. A pair of macromolecules (usually proteins),  $\alpha$  and  $\beta$ , are noncovalently bound in a reaction catalysed by a cognate smg protein, and driven energetically by GTP hydrolysis. A detailed reaction mechanism would be as follows:

- (1)  $\alpha + \text{smg(GTP)} \rightleftharpoons \alpha \cdot \text{smg(GTP)}$
- (2)  $\alpha \cdot \text{smg(GTP)} + \beta \rightleftharpoons \alpha \cdot \beta \cdot \text{smg(GTP)}$
- (3)  $\alpha \cdot \beta \cdot \text{smg(GTP)} \rightleftharpoons \alpha \cdot \beta \cdot \text{smg(GDP)} + \text{P}_i$
- (4)  $\alpha \cdot \beta \cdot \text{smg(GDP)} \rightleftharpoons \alpha \cdot \beta + \text{smg(GDP)}$
- (5)  $\text{smg(GDP)} + \text{GTP} \rightleftharpoons \text{smg(GTP)} + \text{GDP}$

The overall reaction is  $\alpha + \beta + \text{GTP} \rightleftharpoons \alpha \cdot \beta + \text{GDP} + \text{P}_i$ . As smg(GTP) is regenerated (step 5), it is formally a catalyst. The regeneration of smg(GTP) and the hydrolysis of GTP (step 3) can be regulated or catalysed by other proteins such as GAPs and nucleotide-exchanging enzymes and inhibitors.



the course of a cell's life span. Specificity is thus only assured in part by the specific docking of a transport vesicle to its target membrane. Specificity also requires suppressing 'dark reactions' by keeping the fusion machinery effectively 'off' before targeting, and 'off' again the moment fusion is completed. Controlled assembly-disassembly of the NSF-containing fusion particle seems to be at the heart of this crucial regulation and recycling. Assembly is most probably completed from dispersed (that is inactive) subunits only during or after vesicle targeting, and the particle is dismembered as an intrinsic part of the ATP-hydrolysis-driven stroke of NSF that drives fusion (Fig. 4).

It is not yet clear whether additional (as yet unknown) subunits are required for the 20S core particle to catalyse fusion, an evaluation of which awaits reconstitution studies with liposomes, although this seems likely. In particular an smg protein, rab1p (termed ypt in yeast), is needed for attachment/fusion of ER-derived vesicles to the Golgi<sup>80,81</sup> and also for transport in the Golgi stack<sup>82</sup>.

How fusion particles are arranged at a vesicle-cisterna junction, and the biophysical mechanism of fusion are a complete mystery. An interesting speculation<sup>83</sup> is that fusion is promoted by the creation of a protein bridge between adjacent bilayer surfaces, a kind of 'gap junction' with an internal aqueous channel lined by a hydrophobic protein surface. This surface would destabilize the bilayers by providing an energetically similar alternative to the bilayer, creating a hydrophobic pathway which would allow the lipids of the apposed cytoplasmic monolayers to diffuse laterally across the gap until they join, triggering fusion. The model is attractive because it could explain how essentially soluble proteins such as NSF and SNAPs could assemble at the interface of two membranes and promote fusion without actually having to penetrate a lipid bilayer. Conceivably the amphipathic nature of one or more subunits of the fusion particle (such as  $\gamma$ -SNAP) could be exploited to create transiently a hydrophobic channel in or between fusion particles spanning the gap between vesicle and target, the necessary quaternary rearrangements perhaps driven by the hydrolysis of ATP by NSF.

### Generality of the Golgi model

Studies with *S. cerevisiae* provide a satisfying genetic counterpoint to the picture of budding and fusion that has emerged from the biochemical studies with the animal Golgi cell-free system. Temperature-sensitive mutations in any one of three *sec* genes (*sec17*, *sec18* and *sec22*) block transport from ER to Golgi and accumulate 50-nm vesicles in the cytoplasm (Table 3), likely to be transport vesicles derived from the ER<sup>84</sup>. The *sec18* gene encodes the yeast homologue of NSF, a protein that can even replace NSF in the animal cell Golgi transport reaction<sup>73</sup>. The *sec18* protein is also needed for endocytosis in yeast<sup>85</sup>, and recent studies have shown that *sec18* gene function is needed for intraGolgi and Golgi to cell-surface transport in yeast<sup>86</sup>, though perhaps not for Golgi to vacuole transport.

This emphasizes the generality of the NSF-dependent fusion pathway, also shown to be required in a variety of animal cell-free systems for fusion at multiple levels of the Golgi stack<sup>43,87</sup>, for fusion of ER-derived vesicles with the Golgi<sup>88</sup>, and for endosome-endosome fusion<sup>89</sup>. That a general NSF-dependent fusion machinery can be used at so many compartments without a loss of specificity is no longer surprising; once targeting has occurred in a compartment-specific fashion, a general fusion pathway can be called into action without loss of specificity. Not all intracellular fusion events seem to require NSF<sup>86,90</sup>, so there may be other perhaps related molecules in cells.

The *sec17* protein is the functional equivalent in yeast of mammalian  $\alpha$ -SNAP<sup>76</sup>, fitting well with the genetic interactions noted between *sec17* and *sec18* mutations<sup>84</sup>. Like *sec18*, *sec17* gene function is needed throughout the secretory pathway in yeast, confirming that  $\alpha$ -SNAP is used to attach NSF to multiple

TABLE 3 Corresponding proteins and genes implicated in attachment and fusion of uncoated transport vesicles in animals and yeast

| Animal cells         | Yeast cells         |
|----------------------|---------------------|
| NSF (76K)            | <i>sec18p</i> (82K) |
| $\alpha$ -SNAP (35K) | <i>sec17p</i> (34K) |
| $\beta$ -SNAP (36K)  | —                   |
| $\gamma$ -SNAP (39K) | —                   |
| SNAP receptor        | —                   |
| rab1p (23K)          | ypt1p (23K)         |
| —                    | <i>sec22</i> gene   |

compartments for fusion<sup>86</sup>. The role of *sec22* in fusion remains to be clarified.

Mutations in a distinct set of yeast *sec* genes (*sec12*, -13, -16, -21 and -23) also disrupt transport between ER and Golgi, but do not accumulate 50-nm vesicles<sup>84</sup>. Epistasis tests<sup>84</sup> confirm that these 'budding' genes (Table 1) act earlier in the transport pathway than the fusion genes (*sec17*, -18 and -22). Consistent with these assignments, a functionally active transport-vesicle intermediate is produced from ER in crude yeast lysates that can fuse with Golgi *in vitro*<sup>81,91</sup>. Production of the transport intermediate *in vitro* requires a number of budding genes as well as the smg protein, sar1p (refs 63, 81); consumption of the intermediate requires fusion gene products, as well as the smg protein, rab1/ypt (refs 80, 81). The *sec12* gene encodes an integral membrane protein required for sar1p to bind to membranes<sup>63</sup>. Transport vesicles accumulating in *sec18* mutant extracts are unattached to membranes<sup>81</sup>, indicating that NSF is needed to maintain stable association of uncoated vesicles with target membranes, in addition to its role in fusion.

At least two budding *sec* gene products (*sec13p* and *sec23p*) are present in cytosol in high- $M_r$  complexes. The *sec23p* (85K) is bound to a protein of 105K<sup>92</sup>, and *sec13p* (a 34K protein) is in a complex of about 400K<sup>93</sup> that has not been purified. It seems likely that at least some of the budding gene products will prove to be components of coats of the yeast transport vesicles, and constitute subunits of the yeast equivalent of the animal coatamer. The limitations of electron microscopy with yeast make it impossible to discern whether the 50-nm vesicles seen *in vivo* are coated, although as they accumulate with mutations in *sec18* (NSF) and *sec17* gene products (SNAP) it would be predicted (Fig. 2) that they would already have lost their coats.

How general are COP-coated vesicles as carriers of the bulk flow to the cell surface? Vesicles with this morphology are seen budding from transitional elements of ER and the TGN<sup>94</sup>, and coatamer (as judged by its  $\beta$ -COP subunit) accumulates at both the CGN and the TGN during appropriate temperature blocks that retard exit from these compartments<sup>58</sup>. Thus a case can be made for COP-coated vesicles being responsible for every step

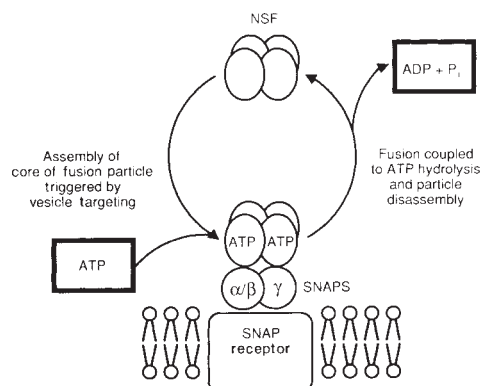


FIG. 4 Regulation of membrane fusion by controlled subunit assembly, driven by ATP hydrolysis. Each of the steps involving NSF, SNAPs and SNAP receptor has been demonstrated (see text). The involvement of additional proteins in the fusion cycle seems likely.

of anterograde transport from the ER to the plasma membrane. Consistent with this possibility are striking similarities of ER to Golgi cell-free systems from yeast and animal systems<sup>81,91,95-97</sup> to the Golgi system. GTP is required for ER vesicle budding<sup>81</sup> and secretory storage vesicle budding from TGN (ref. 98). Post-Golgi vesicles from yeast have been purified and are not coated<sup>99</sup>. Also, vesicles released from the TGN and targeted to apical or basolateral plasma membrane domains<sup>100</sup> and transcytotic vesicles<sup>101</sup> have all been purified, and all seem to lack coats. It is at present impossible to determine whether any or all of these kinds of vesicles originally formed with coats that were then shed.

## Summary

Although important questions concerning the cell biology of compartmental organization of the secretory pathway still remain, enough is now known to permit a meaningful dissection of the molecular machinery of individual segments. It is both fortunate and remarkable that these kinds of pathways, whose essential purpose is to propagate the three-dimensional organization of the cytoplasm, can nonetheless be faithfully reproduced in dispersed cell-free systems without the benefit (or constraint) of pre-existing spatial arrangements. This has opened the door to biochemistry, and superficial outlines of the steps involved in vesicle budding and fusion are now emerging. Crucial points remain. Among them, how in step-by-step fashion do coats assemble on membranes to yield a vesicle? What is the targeting

signal that triggers uncoating and attachment of a vesicle? How can a protein machine fuse two lipid layers?

Answers to questions of molecular mechanism at this level have and will necessarily continue to emerge from cell-free systems with allied methods of morphology, and thus need in one way or another to be confirmed *in vivo*. As so little is known or can be learned at this level from studies in whole cells, how can this be done? The answer is that the molecules discovered with cell-free systems, putatively performing the same roles in living cells, will provide the very tools to make the assessment of authenticity. As genes encoding these purified transport components are manipulated, and antibodies microinjected, the predicted effects on cellular physiology can be scrutinized. This has already begun by synthesizing the results from animal cell-free biochemistry with those from yeast genetics, and the results are encouraging. We can now look forward to a rapid increase in our knowledge of the secretory pathway so that soon we will be able to discuss these complex events of macromolecular targeting with the same kind of sophistication with which we now describe the biosynthesis of macromolecules such as proteins and nucleic acids, and will be able to do so in a common language, that of protein biochemistry. □

James E. Rothman is at the Program of Cellular Biochemistry and Biophysics, Rockefeller Research Laboratory, Sloan-Kettering Institute, 1275 York Avenue, New York, New York 10021, USA; Lelio Orci is at the Department of Morphology, Institute of Histology and Embryology, University of Geneva Medical School, 1 rue Michel Servet, 1211 Geneva 4, Switzerland.

- Palade, G. E. *Science* **189**, 347-358 (1975).
- Fries, E. & Rothman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3870-3814 (1980).
- Novick, P., Fero, S. & Schekman, R. *Cell* **25**, 461-469 (1981).
- Griffiths, G. & Simons, K. *Science* **234**, 438-443 (1986).
- Rothman, J. E., Miller, R. L. & Urbani, L. J. *J. Cell Biol.* **99**, 260-371 (1984).
- Dunphy, W. G. & Rothman, J. E. *Cell* **42**, 13-21 (1985).
- Kornfeld, R. & Kornfeld, S. A. *Rev. Biochem.* **54**, 631-664 (1985).
- Farguhar, M. G. A. *Rev. Cell Biol.* **1**, 447-488 (1985).
- Saraste, J. & Kuismunen, E. *Cell* **38**, 535-549 (1984).
- Lippincott-Schwartz, J. *et al. Cell* **60**, 821-836 (1990).
- Huttner, W. B. & Tooze, S. A. *Curr. Opin. Cell Biol.* **1**, 648-654 (1989).
- Hsu, V. *et al. Nature* **352**, 441-444 (1991).
- Mellman, I. & Simons, K. *Cell* (in the press).
- Burgess, T. L. & Kelly, R. B. A. *Rev. Cell Biol.* **3**, 243-293 (1987).
- Rothman, J. E. *Cell* **50**, 521-522 (1987).
- Hurtley, S. M. & Helenius, A. A. *Rev. Cell Biol.* **5**, 277-307 (1989).
- Wieland, F. *et al. Cell* **50**, 289-300 (1987).
- Lodish, H. F. & Kong, N. J. *biol. Chem.* **266**, 14835-14838 (1991).
- Munro, S. & Pelham, H. R. B. *Cell* **48**, 899-907 (1987).
- Nilsson, T., Jackson, M. & Peterson, P. A. *Cell* **58**, 707-718 (1989).
- Swift, A. M. & Machamer, C. E. *J. Cell Biol.* (in the press).
- Dean, H. & Pelham, H. R. B. *J. Cell Biol.* **111**, 369-377 (1990).
- Lewis, M. J., Sweet, D. J. & Pelham, H. R. B. *Cell* **61**, 1359-1363 (1990).
- Semenza, J. C. *et al. Cell* **61**, 1349-1357 (1990).
- Lewis, M. J. & Pelham, H. R. B. *Nature* **348**, 162-163 (1990).
- Vaux, D., Tooze, J. & Fuller, S. *Nature* **345**, 495-502 (1990).
- Lippincott-Schwartz, J. *et al. Cell* **60**, 821-836 (1990).
- Jackson, M., Nilsson, T. & Peterson, P. A. *EMBO J.* **9**, 3153-3162 (1990).
- Pelham, H. R. B. *EMBO J.* **7**, 913-918 (1989).
- Brands, R. *et al. J. Cell Biol.* **101**, 1724-1732 (1985).
- Rothman, J. E. *Science* **213**, 1212-1219 (1981).
- Balch, W. E. *J. biol. Chem.* **264**, 16965-16968 (1989).
- Goda, Y. & Pfeffer, S. R. *FASEB J.* **3**, 2488-2495 (1989).
- Glick, B. S. & Rothman, J. E. *Cell* **39**, 525-536 (1984).
- Orci, L., Glick, B. S. & Rothman, J. E. *Cell* **46**, 171-184 (1986).
- Balch, W. E. *et al. Cell* **39**, 405-416 (1984).
- Dunphy, W. G., Brands, R. & Rothman, J. E. *Cell* **40**, 463-472 (1985).
- Dunphy, W. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 7453-7457 (1981).
- Rothman, J. E., Urbani, L. H. & Brands, R. *J. Cell Biol.* **99**, 248-259 (1984).
- Braell, W. A. *et al. Cell* **39**, 511-524 (1984).
- Melancon, P. *et al. Cell* **51**, 1053-1062 (1987).
- Glick, B. S. & Rothman, J. E. *Nature* **326**, 309-312 (1987).
- Malhotra, V. *et al. Cell* **54**, 221-227 (1988).
- Orci, L. *et al. Cell* **56**, 357-368 (1989).
- Block, M. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7852-7856 (1988).
- Pfanner, N. *et al. Cell* **59**, 95-102 (1989).
- Pfanner, N. *et al. J. Cell Biol.* **110**, 955-961 (1990).
- Goldstein, J. L., Anderson, R. W. G. & Brown, M. S. *Nature* **270**, 679-685 (1979).
- Pearse, B. M. F. & Bretscher, M. S. A. *Rev. Biochem.* **50**, 85-101 (1981).
- Anderson, R. W. G., Brown, M. S. & Goldstein, J. L. *Cell* **10**, 351-364 (1977).
- Rothman, J. E. & Schmid, S. L. *Cell* **46**, 5-9 (1986).
- Duden, R., Allen, V. & Kreis, T. *Trends Cell Biol.* **1**, 14-91 (1991).
- Malhotra, V. *et al. Cell* **58**, 329-336 (1989).
- Serafini, T. *et al. Nature* **349**, 215-220 (1991).
- Serafini, T. *et al. Cell* **67**, 239-253 (1991).
- Kahn, R. & Gilman, A. J. *biol. Chem.* **261**, 7906-7911 (1986).
- Stearns, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 1238-1242 (1990).
- Duden, R. *et al. Cell* **64**, 649-665 (1991).
- Orci, L. *et al. Cell* **64**, 1183-1195 (1991).
- Waters, M. G., Serafini, T. & Rothman, J. E. *Nature* **349**, 248-251 (1991).
- Kahn, R. *et al.*, manuscript submitted.
- Salminen, A. & Novick, P. J. *Cell* **49**, 527-538 (1987).
- d'Enfert, C. *et al. J. Cell Biol.* **114**, 663-670 (1991).
- Kahn, R., Goddard, C. & Newkirk, M. J. *biol. Chem.* **263**, 8282-8287 (1987).
- Kahn, R. *et al. J. biol. Chem.* **266**, 2606-2614 (1991).
- Balch, W. E. *Trends Biochem. Sci.* **15**, 473-477 (1990).
- Chavrier, P. *et al. Nature* **353**, 769-772 (1991).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
- Stow, J. *et al. J. Cell Biol.* **114**, 1113-1124 (1991).
- Donaldson, J. G. *et al. Nature* (in the press).
- Pearse, B. M. F. & Robinson, M. A. *Rev. Cell Biol.* **6**, 151-171 (1990).
- Ahle, S. *et al. EMBO J.* **7**, 919-929 (1988).
- Wilson, D. W. *et al. Nature* **339**, 355-359 (1989).
- Tagaya, M. *et al.*, manuscript submitted.
- Weidman, P. *et al. J. Cell Biol.* **108**, 1589-1596 (1989).
- Clary, D. O., Griff, I. C. & Rothman, J. E. *Cell* **61**, 789-721 (1990).
- Clary, D. O. & Rothman, J. E. *J. biol. Chem.* **265**, 10109-10117 (1990).
- Whitehart, S. W. *et al.*, manuscript submitted.
- Wilson, D. W. *et al.*, manuscript submitted.
- Segue, N. *Science* **252**, 1553-1556 (1991).
- Rexach, M. F. & Schekman, R. W. *J. Cell Biol.* **114**, 219-229 (1991).
- Plutner, H. A. *et al. J. Cell Biol.* **115**, 31-43 (1991).
- Bentz, J. in *Drug and Anesthetic Effects on Membrane Structure and Function* 259-287 (Wiley-Liss, New York, 1991).
- Kaiser, C. A. & Schekman, R. *Cell* **61**, 723-733 (1990).
- Riezman, H. *Cell* **40**, 1001-1009 (1985).
- Graham, T. R. & Emr, S. J. *Cell Biol.* **114**, 207-218 (1991).
- Rothman, J. E. *J. biol. Chem.* **265**, 12502-12510 (1987).
- Beckers, C. J. M. *et al. Nature* **339**, 397-398 (1989).
- Diaz, R. *et al. Nature* **339**, 398-400 (1989).
- Goda, Y. & Pfeffer, S. R. *J. Cell Biol.* **112**, 823-831 (1991).
- Groesch, M. E. *et al. J. Cell Biol.* **111**, 45-53 (1990).
- Schekman, R. *et al. J. Cell Biol.* **109**, 2a (1989).
- Pryer, N. K. *et al. J. Cell Biol.* **111**, 325a (1990).
- Griffiths, G. *et al. J. Cell Biol.* **101**, 949-964 (1985).
- Baker, D. *et al. Cell* **54**, 335-344 (1988).
- Ruohola, H., Kabaceni, A. K. & Ferro-Novick, S. J. *Cell Biol.* **107**, 1465-1476 (1988).
- Beckers, C. J. M., Keller, D. S. & Balch, W. E. *Cell* **50**, 523-534 (1987).
- Tooze, S. A. & Huttner, W. B. *Cell* **60**, 837-847 (1990).
- Walworth, N. C. & Novick, P. J. *J. Cell Biol.* **105**, 163-174 (1987).
- deCurtis, I. & Simons, K. *Cell* **58**, 719-727 (1989).
- Stzul, E., Kaplan, A., Saucan, L. & Palade, G. *Cell* **64**, 81-89 (1991).
- Haubruck, H. *et al. EMBO J.* **8**, 1427-1432 (1989).
- Chavrier, P. *et al. Cell* **62**, 317-329 (1990).
- von Mollard, G. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 1988-1992 (1990).
- Darchen, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 5692-5696 (1990).
- Goud, B. & McCaffrey, M. *Curr. Opin. Cell Biol.* **3**, 626-633 (1991).
- Gorvel, J. P. *et al. Cell* **64**, 915-925 (1991).
- Goud, B. *et al. Nature* **345**, 533-556 (1990).
- Goud, B. *et al. Cell* **53**, 753-768 (1988).
- Zahraoui, A. *et al. J. biol. Chem.* **264**, 12394-12401 (1989).
- Chavrier, P. *et al. Molec. cell Biol.* **10**, 6578-6585 (1990).

ACKNOWLEDGEMENTS. We thank J. Lenard, P. Marks, I. Mellman, P. Pederson, S. Pfeffer, R. Rifkin, K. Simons, G. Warren and F. Wieland for their comments. Work from the authors' laboratories was supported by the US NIH, the Swiss National Science Foundation and the Human Frontiers Science Program.



# Evidence from Southern Ocean sediments for the effect of North Atlantic deep-water flux on climate

Christopher D. Charles & Richard G. Fairbanks

Lamont-Doherty Geological Observatory, Palisades, New York, 10964, USA  
Department of Geology, Columbia University, New York, New York 10025, USA

The Southern Ocean is perhaps the only region where fluctuations in the global influence of North Atlantic Deep Water (NADW) can be monitored unambiguously in single deep-sea cores. A carbon isotope record from benthic foraminifera in a Southern Ocean core reveals large and rapid changes in the flux of NADW during the last deglaciation, and an abrupt increase in the NADW production rate which immediately preceded large-scale melting of the Northern Hemisphere ice sheets. This sudden strengthening of the NADW thermohaline cell provides strong evidence for the importance of NADW in glacial-interglacial climate change.

FUNDAMENTAL aspects of the Earth's climate variability over the last deglaciation indicate the operation of strong internal feedbacks and amplifiers in the global climate system. The rapidity and amplitude of the glacial-interglacial transition<sup>1-4</sup>, the pulsed destruction of the Northern Hemisphere ice sheets<sup>4</sup>, and the interhemispheric synchronicity of change<sup>5,6</sup> are all difficult to explain by solar insolation changes alone. Two mechanisms often invoked to explain these phenomena are variations in greenhouse gas concentrations and poleward oceanic heat transport (NADW formation)<sup>2,7-11</sup>. The basic principles of how both these amplifiers might work are well established (see ref. 12 for a review and further references), yet rigorous geological assessment of how either was actually involved with the last deglaciation has not been possible.

Precise radiocarbon dating of the Barbados sea-level record, which is marked by two intervals of massive ice-sheet melting centred on 12.0 kyr and 9.5 kyr (ref. 4), now provides an explicit geological test for these proposed climate amplifiers. Comparable chronologies for atmospheric and deep-ocean variability over the last deglaciation are clearly required. Existing deglacial ice-core records of greenhouse gases suffer either from summer melting artefacts<sup>13</sup> or dating uncertainties<sup>14</sup> and are therefore not appropriate for comparison to the sea-level record. Similarly, no clear consensus on the timing of deep-water variability has emerged<sup>4,12,15,16</sup> and the evidence for deglacial deep-water fluctuations has thus far been limited to North Atlantic records<sup>15,16</sup> which could be overprinted by local deep-water mixing effects<sup>16,17</sup>.

Our approach to this problem has been to focus on Southern Ocean deep-sea sediment cores. The nutrient chemistry of the Southern Ocean is strongly influenced by the influx of NADW, and the global-scale mixing of deep waters that occurs in the Antarctic circumpolar current provides a filter for local variability. Thus, Southern Ocean palaeonutrient records can, in principle, provide a suitable deep-ocean counterpart to the Barbados

coral record of ice-sheet decay. Here we present a radiocarbon-dated  $\delta^{13}\text{C}$  record from benthic foraminifera in a Southern Ocean deep-sea core, with resolution detailed enough to test the interactions between the Northern Hemisphere ice sheets and deep-ocean circulation over the last deglaciation.

## NADW and circumpolar $\delta^{13}\text{C}$

Figure 1 illustrates why the  $\delta^{13}\text{C}$  of circumpolar water is sensitive to the variations in the contribution of NADW. Circumpolar Deep Water (CPDW) represents a mixture of NADW and recirculated deep water from the Indian and Pacific. Consequently, CPDW  $\delta^{13}\text{C}$  values lie between those of NADW (characterized by high  $\delta^{13}\text{C}$  and low nutrients) and Indo-Pacific Deep Water (low  $\delta^{13}\text{C}$  and high nutrients). Without any NADW influence, the  $\delta^{13}\text{C}$  of CPDW should decrease to Indo-Pacific values, regardless of how the oceanic inventory of  $^{13}\text{C}$  or  $^{12}\text{C}$  might have changed. Comparison of North Atlantic, Pacific and Southern Ocean  $\delta^{13}\text{C}$  records from benthic foraminifera confirmed this concept<sup>17</sup>, highlighting a primary advantage of using CPDW  $\delta^{13}\text{C}$  records for documenting the history of NADW: any uncertainties involving the mixing ratio of northern and southern sources are circumvented. On the other hand, the problem with using North Atlantic  $\delta^{13}\text{C}$  records is that the measured changes may simply reflect shifts in the mixing zone between Antarctic Bottom Water (AABW) and NADW, with no relationship to net NADW flux. Thus, although it may be less sensitive to subtle changes in NADW strength because it is 'downstream' of the source, the Southern Ocean is perhaps the only region where the global influence of NADW fluctuations can be monitored unambiguously in single locations. This strategic advantage is important, because deep-sea cores will never begin to capture the three-dimensional structure of the deep ocean in the past.

It is possible, however, that deviations from simple mixing models for CPDW  $\delta^{13}\text{C}$  can occur. For example, there is growing evidence that  $\delta^{13}\text{C}$  in Southern Ocean foraminifera was lower (by  $\sim 0.2\%$ ) than in those from the deep Pacific during the glacial periods<sup>18</sup>, and this distribution of  $\delta^{13}\text{C}$  cannot be easily explained by circulation alone. Nonconservative effects involving the ecology of benthic foraminifera<sup>19</sup>, the varying degrees of isotopic equilibrium reached by incorporated surface waters<sup>20</sup> and the degradation of specific organic matter<sup>21</sup> are all possible. Although the first of these effects must be assessed empirically, mass balance considerations suggest that the latter two effects are small. As the glacial and interglacial amplitude of foraminiferal  $\delta^{13}\text{C}$  in Southern Ocean cores is  $\sim 1\%$  (as is the present Atlantic-Pacific  $\delta^{13}\text{C}$  gradient in total inorganic carbon), nonconservative effects of 0.2% do not preclude using  $\delta^{13}\text{C}$  to deduce large-scale changes in water mass circulation. A more serious challenge is the observation that Cd/Ca ratios in benthic foraminifera (another palaeonutrient proxy) from Southern Ocean cores show little glacial-interglacial change<sup>21</sup>. We have no definitive explanation for this apparent discrepancy, although recent data indicate that foraminiferal distribution coefficients for Cd may vary with depth by a factor of two<sup>22</sup>. In any case,

we make the simplest interpretation that seems justified from the modern distribution of  $\delta^{13}\text{C}$  in deep waters and foraminifera: a strong relative contribution of NADW is the only means for increasing CPDW  $\delta^{13}\text{C}$  above Pacific (mean ocean) values.

### Changes over the last deglaciation

Core RC11-83, recovered on the north flank of the Cape rise (41° 36' S, 94° 48' E; 4,718 m) was overlain by deep waters indistinguishable from Lower Circumpolar Deep Water. The core is characterized by sedimentation rates of roughly 25 cm kyr<sup>-1</sup>; consequently, its benthic records spanning the past 15 kyr represent as yet the most detailed chronicle of deglacial isotopic change in the deep Southern Ocean. We generated stable isotope records for *Planulina wuellerstorfi*, whose shells seem to be relatively reliable recorders of the isotopic composition of the ambient water<sup>18,19</sup>.

The deglacial timescale for this record is derived from accelerator mass spectrometric (AMS) <sup>14</sup>C dates on seven levels in the core (listed in Table 1; ages given throughout the text refer to <sup>14</sup>C kyr BP) and a polynomial fit of the resulting age-depth curve. The ages obtained from *Globigerina bulloides* have been corrected by a constant 600 yr, to account for the fact that the waters in which these foraminifera grow are not in isotopic equilibrium with the atmosphere. This constant correction is likely to be a reasonable estimate of the reservoir age (within 200 yrs) throughout deglaciation<sup>23</sup>. The  $\delta^{18}\text{O}$  time series derived from this age model (not shown) is in excellent agreement with previously published deglacial  $\delta^{18}\text{O}$  curves from the Southern Ocean<sup>24</sup>.

The RC11-83 benthic  $\delta^{13}\text{C}$  record (Figs 1 and 2) demonstrates that deglacial changes in the nutrient chemistry of the deep Southern Ocean were rapid and of considerable magnitude. At 16 kyr, the  $\delta^{13}\text{C}$  values are the same as<sup>25</sup>, or slightly lower than<sup>18,26</sup> analogous glacial deep Pacific values, confirming earlier findings of the reduced influence of NADW in the Southern Ocean during the last glacial maximum<sup>17</sup>. But an abrupt shift from low values characteristic of glacial conditions to high, essentially modern, values occurs from 12.6–12.2 kyr. Following the premise described in the previous section, this rapid rise in Southern Ocean  $\delta^{13}\text{C}$  values must have been associated with the 'turn on' of NADW. As the duration of this abrupt  $\delta^{13}\text{C}$  transition (400 years) is of the same order as the mixing time of the Atlantic ocean<sup>27</sup>, the change from low to high relative NADW flux appears virtually instantaneous from a geological standpoint. The RC11-83 record also shows significant oscillations superimposed on this basic glacial-interglacial change. An excursion towards higher values occurs between 13.1 and 12.7 kyr, and a low- $\delta^{13}\text{C}$  event occurs at 11.5 kyr.

Comparison with another detailed deglacial record<sup>16</sup> from the deep North Atlantic (Fig. 1) indicates that  $\delta^{13}\text{C}$  changes in RC11-83 can be attributed directly to the varying contribution of NADW in the Southern Ocean. The striking transition from nutrient enrichment (<sup>13</sup>C depletion) to nutrient depletion (<sup>13</sup>C enrichment), which is dated between 12.6 and 12.2 kyr in RC11-83, is also obvious in the North Atlantic core. Another step in the same direction, but with lower amplitude, is apparent from ~10–9 kyr. In addition, the North Atlantic record shows an anomaly similar to, and perhaps correlated with, the excursion

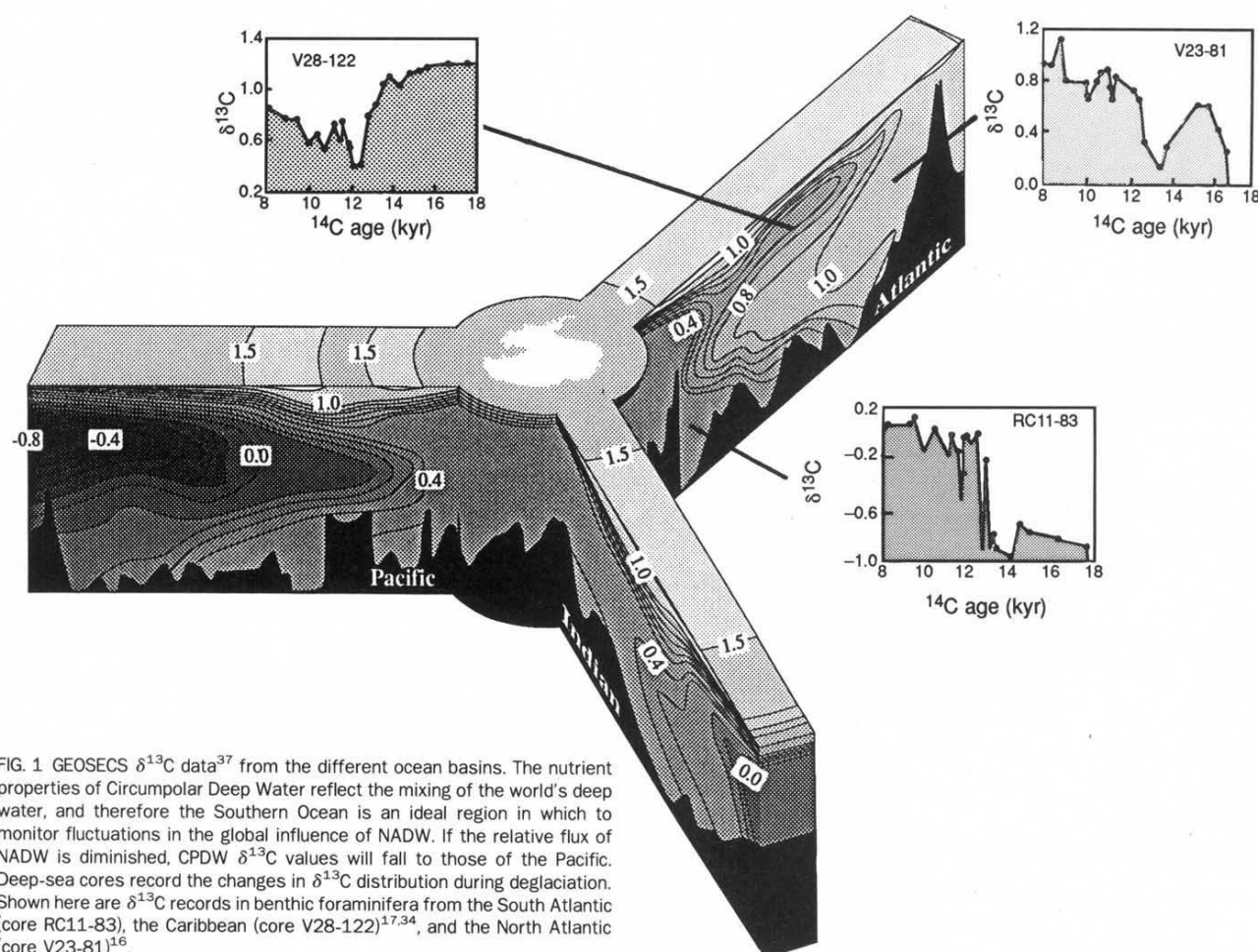


FIG. 1 GEOSECS  $\delta^{13}\text{C}$  data<sup>37</sup> from the different ocean basins. The nutrient properties of Circumpolar Deep Water reflect the mixing of the world's deep water, and therefore the Southern Ocean is an ideal region in which to monitor fluctuations in the global influence of NADW. If the relative flux of NADW is diminished, CPDW  $\delta^{13}\text{C}$  values will fall to those of the Pacific. Deep-sea cores record the changes in  $\delta^{13}\text{C}$  distribution during deglaciation. Shown here are  $\delta^{13}\text{C}$  records in benthic foraminifera from the South Atlantic (core RC11-83), the Caribbean (core V28-122)<sup>47,34</sup>, and the North Atlantic (core V23-81)<sup>16</sup>.



TABLE 1 Data for core RC11-83

| Depth<br>(cm) | Age<br>( <sup>14</sup> C kyr) | Stable isotopes<br>(Lamont-Doherty<br>corrected*) |                                  | Accelerator radiocarbon ages<br>(NSF-Arizona accelerator<br>facility, Tucson) |                            |
|---------------|-------------------------------|---------------------------------------------------|----------------------------------|-------------------------------------------------------------------------------|----------------------------|
|               |                               | $\delta^{18}\text{O}$<br>(‰ PDB)                  | $\delta^{13}\text{C}$<br>(‰ PDB) | Accession<br>number                                                           | Age<br><i>G. bulloides</i> |
| 165           | 7.57                          | 2.39                                              | -0.03                            |                                                                               |                            |
| 170           | 7.92                          | 2.41                                              | 0.07                             |                                                                               |                            |
| 175           | 8.27                          | 2.64                                              | 0.13                             |                                                                               |                            |
| 189           | 9.38                          | 2.66                                              | 0.14                             | 5,974                                                                         | 9.90 ± 0.066               |
| 189           |                               |                                                   |                                  | 5,974                                                                         | 10.08 ± 0.090              |
| 191           | 9.53                          | 2.60                                              | 0.19                             |                                                                               |                            |
| 197           | 9.97                          | 2.54                                              | -0.07                            |                                                                               |                            |
| 205           | 10.48                         | 2.75                                              | 0.10                             |                                                                               |                            |
| 215           | 11.03                         | 3.09                                              | -0.09                            |                                                                               |                            |
| 217           | 11.12                         | 2.94                                              | -0.11                            |                                                                               |                            |
| 220           | 11.27                         | 2.64                                              | 0.06                             |                                                                               |                            |
| 221           |                               |                                                   |                                  | 5,975                                                                         | 11.94 ± 0.08               |
| 221           |                               |                                                   |                                  | 5,975                                                                         | 11.93 ± 0.10               |
| 225           | 11.49                         | 2.71                                              | -0.08                            |                                                                               |                            |
| 228           | 11.61                         | 3.02                                              | -0.07                            |                                                                               |                            |
| 231           |                               |                                                   |                                  | 5,976                                                                         | 12.04 ± 0.08               |
| 231           |                               |                                                   |                                  | 5,976                                                                         | 12.19 ± 0.12               |
| 232           | 11.77                         | 3.24                                              | -0.47                            |                                                                               |                            |
| 233           | 11.80                         | 3.22                                              | -0.25                            |                                                                               |                            |
| 235           | 11.87                         | 3.26                                              | 0.03                             |                                                                               |                            |
| 238           | 11.98                         | 3.35                                              | 0.04                             |                                                                               |                            |
| 245           | 12.20                         | 3.38                                              | -0.02                            |                                                                               |                            |
| 248           |                               |                                                   |                                  | 5,977                                                                         | 12.79 ± 0.11               |
| 248           |                               |                                                   |                                  | 5,977                                                                         | 13.02 ± 0.09               |
| 255           | 12.49                         | 3.44                                              | 0.07                             |                                                                               |                            |
| 260           | 12.62                         | 3.35                                              | -0.62                            |                                                                               |                            |
| 262           | 12.67                         |                                                   |                                  | 5,978                                                                         | 13.23 ± 0.11               |
| 265           | 12.75                         | 3.38                                              | -0.59                            |                                                                               |                            |
| 270           | 12.87                         | 3.29                                              | -0.86                            |                                                                               |                            |
| 275           | 13.00                         | 3.44                                              | -0.14                            |                                                                               |                            |
| 278           | 13.07                         | 3.56                                              | -0.30                            |                                                                               |                            |
| 285           | 13.24                         | 3.76                                              | -0.85                            |                                                                               |                            |
| 288           | 13.32                         | 3.60                                              | -0.72                            |                                                                               |                            |
| 295           | 13.50                         | 3.81                                              | -0.83                            |                                                                               |                            |
| 298           | 13.59                         |                                                   |                                  | 5,979                                                                         | 14.31 ± 0.11               |
| 315           | 14.13                         | 3.81                                              | -0.91                            |                                                                               |                            |
| 325           | 14.53                         | 3.87                                              | -0.65                            |                                                                               |                            |
| 335           | 15.00                         | 3.93                                              | -0.70                            | 5,980                                                                         | 15.62 ± 0.12               |
| 365           | 17.02                         | 3.91                                              | -0.74                            |                                                                               |                            |
| 395           | 20.24                         | 3.98                                              | -0.81                            |                                                                               |                            |

\* Age adjusted by 600 yr to correct for apparent age of surface seawater<sup>23</sup>.  $\delta^{18}\text{O} = [({}^{18}\text{O}/{}^{16}\text{O})_{\text{sample}}/({}^{18}\text{O}/{}^{16}\text{O})_{\text{standard}}] - 1$  in ‰;  $\delta^{13}\text{C} = [({}^{13}\text{C}/{}^{12}\text{C})_{\text{sample}}/({}^{13}\text{C}/{}^{12}\text{C})_{\text{standard}}] - 1$ .

at 11.5 kyr in RC11-83. The records also agree in the fact that the Younger Dryas interval (11–10 kyr) is not characterized by a particularly strong anomaly. Any similarities observed in both the North and South Atlantic must reflect the varying influence of NADW during deglaciation because the southward flow of nutrient-depleted NADW repressed the northward flow of nutrient-rich Southern Ocean water to the deep North Atlantic.

Where there is disagreement between the various nutrient proxy records for the Atlantic, we argue that the Southern Ocean would yield the more comprehensive measure of NADW flux variability. This issue is important for the interval from 16 to 13.5 kyr, because the North Atlantic record shows a deep minimum centred on 13.5 kyr, suggesting that the Last Glacial Maximum was not the time of lowest NADW production. The RC11-83 record, in contrast, shows more uniformly low values until 13 kyr, suggesting minimal NADW flux throughout the glacial period. Resolving such differences of interpretation is obviously crucial for an accurate evaluation of NADW's influence in global climate.

Another source of evidence for a profound, rapid deglacial change in Atlantic thermohaline circulation is provided by  $\delta^{13}\text{C}$  records from Caribbean deep-sea cores, which monitor mid-depth North Atlantic water (1,800 m)<sup>28</sup>. In these records, however, the sense of change is opposite to that of the deep Atlantic<sup>15,17</sup>. Although relative nutrient depletion prevailed during glacial periods, a shift to more nutrient-rich conditions occurred at 12.6 kyr, precisely the time of deep-ocean change (Figs. 1 and 2). The inverse mid-depth and deep Atlantic signals could both reflect the same phenomenon if relatively nutrient-rich water from the Southern Ocean were drawn northwards across the Equator at mid-depths by the removal of surface and thermocline water to form southward-flowing NADW. This possibility is completely compatible with Boyle and Keigwin's original explanation for the Caribbean signal, that NADW resided at shallower depths during glacial periods<sup>16</sup>, and with the glacial Atlantic reconstruction of Duplessy *et al.*<sup>26</sup>. In any case, the chemistry of the Atlantic over a large depth range was apparently affected by the abrupt deglacial changes in the NADW cell.

Figure 2 shows that the 'turn on' of NADW deduced from the convergence of the Atlantic nutrient records was nearly synchronous with, or immediately preceded, the first significant pulse of Northern Hemisphere ice-sheet melting measured by the Barbados sea-level curve. The ocean circulation changes also seem to be concomitant with the first signs of warmth in European continental pollen assemblages from lake sediments<sup>29</sup>.

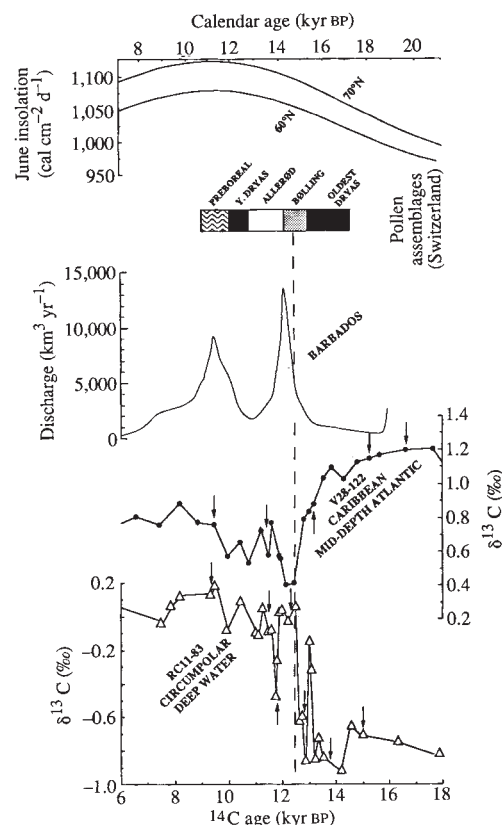


FIG. 2 Convergence of deep Southern Ocean and Caribbean<sup>17,34</sup>  $\delta^{13}\text{C}$  records, with AMS  $^{14}\text{C}$  dated levels indicated by arrows, marks the 'turn on' of NADW (dashed line). The timing of this event can be compared with the ice-sheet meltwater discharge curve (calculated as the first derivative of the Barbados sea-level curve<sup>4,38</sup>), the change in pollen assemblages in Swiss lake sediments (schematic zones taken from ref. 29), and June insolation at 60° N and 70° N (ref. 38), provided that a common timescale is established. The top scale is calendar years, and the bottom scale is  $^{14}\text{C}$  years. Although the ocean records and lake sediment records are dated by  $^{14}\text{C}$ , they can be compared to records dated in calendar years (insolation) by using the calibration in ref. 30.

The records of the different climate variables, correlated on the basis of detailed AMS  $^{14}\text{C}$  dating, may be placed in the context of solar insolation changes through the Barbados coral Th/U calibration of the radiocarbon timescale<sup>30</sup>.

## Discussion

Although there is no simple astronomical explanation for the shape of the glacial meltwater curve<sup>4</sup>, coincidence between the abrupt ocean circulation changes and the pulses of ice-sheet melting strongly suggests that NADW variability was an important factor in the shift to a deglaciated state. The Southern Ocean isotopic record might be taken as evidence for the suggestion that changes in the NADW cell promoted a rapid, global reorganization of the climate system to an interglacial mode of operation<sup>12</sup>. But although the renewal of NADW may seem a simple and chronologically consistent explanation for the meltwater pulses, the timing alone does not constitute proof of this hypothesis.

The questions to be asked are, first, what might have caused the NADW flux changes, and second, whether the heat transport associated with renewed NADW could have provided the trigger for ice-sheet melting. Broecker *et al.* suggest that when Northern Hemisphere ice sheets grew to their maximum extent, the strength of the NADW cell was poised to oscillate, because the salt balance of the North Atlantic involved a precarious interplay of glacial meltwater, ocean advection and water vapour loss<sup>31</sup>. The excursions present in the RC11-83  $\delta^{13}\text{C}$  record from 12.6 to 11.5 kyr might conform to the model of ref. 31: NADW 'turns on', causing ice to melt, which in turn causes NADW production to slow. But the oscillation from 13.1 to 12.6 kyr in the RC11-83 record, which may be considered a 'false start' towards prolonged deep-water formation, is not well resolved in the meltwater curve, and for the salt oscillator model to apply, another salt pathway (such as sea-ice growth in the Norwegian-Greenland or Labrador Seas, or oceanic advection) must have played a stronger part at this time. Therefore, at face value, the RC11-83 record allows for some independence between the deep ocean and the ice sheets.

The immediate effect of North Atlantic heat transports on ice-sheet melting is not entirely obvious. General circulation models clearly demonstrate that continental Europe would be the main beneficiary of any increased oceanic heat drawn poleward by NADW formation<sup>32</sup>. Thus, the correlation between

changes in European pollen assemblage and NADW flux changes measured in the Southern Ocean is not surprising (although the Younger Dryas interval, where the Southern Ocean record shows little change, is an example of how the coupling between the two was not always exact). Yet deep-sea core evidence<sup>33,34</sup> and the mapping of meltwater plumes<sup>35</sup> suggest that the Laurentide (North American) ice sheet was the primary contributor to the first main meltwater pulse. If NADW formation brought warmer water to the Labrador Sea and Baffin Bay, melting along the northeast margin of the Laurentide could be explained, but the southern margin melting poses a problem for this mechanism and perhaps should be a subject for further modelling.

Of course, the 'turn on' of NADW could indirectly affect the global heat budget by altering atmospheric  $\text{CO}_2$  concentrations. The massive redistribution of nutrients at 12.6 kyr, as evidenced by the Atlantic records, must have affected the ocean's capacity to hold  $\text{CO}_2$ , though the deglacial rate of  $\text{CO}_2$  increase is still a matter for debate. The abrupt shift in NADW production, with a consequent increase in heat transport, may actually be responsible for the first deglacial rise in the  $\text{CO}_2$  record in the Greenland ice cores, which has been reinterpreted as an artefact of summer melting<sup>13</sup>. New results from the Byrd core (Antarctica)<sup>13</sup> suggest that the  $\text{CO}_2$  response to ocean circulation and nutrient changes was gradual and thus could not have been the immediate cause of the meltwater pulses. Until an unequivocal  $\text{CO}_2$  record is produced, we prefer to emphasize the possible direct consequences of NADW formation on global climate change.

The Southern Ocean  $\delta^{13}\text{C}$  record therefore demonstrates that large, rapid fluctuations in the production of NADW were felt outside the North Atlantic basin over the last deglaciation. The detailed dating of this record refines previous estimates<sup>12</sup> of 14 kyr for the re-initiation time of NADW to 12.6–12.2 kyr, an interval virtually coincident with the first melting episode of the Northern Hemisphere ice sheets<sup>4</sup>. The record thus provides strong evidence that NADW could have been the primary amplifier of glacial-interglacial climate change in high latitudes. Enigmatic low- to mid-latitude climate phenomena, such as the synchronous shifts of mountain snowlines and floral assemblages in both hemispheres, are left to be explained by other means, possibly  $\text{CO}_2$ -induced changes in plant survival or local climate patterns<sup>36</sup>. □

Received 23 July; accepted 17 December 1991.

- Broecker, W. S. & van Donk, J. *Rev. Geophys. Space Phys.* **8**, 167–198 (1970).
- Broecker, W. S. in *Milankovich and Climate* (eds. Berger, A. *et al.*) 687–698 (Reidel, Dordrecht, 1984).
- Broecker, W. S., Peteet, D. M. & Rind, D. *Nature* **315**, 21–25 (1985).
- Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
- Mercer, J. H. in *Climate Processes and Climate Sensitivity*, AGU Geophys. Monogr. Ser. Vol. 29 (eds Hansen, J. E. & Takahashi, T.) 307–313 (1984).
- Nelson, C. S., Hendy, C. H., Jarrett, G. R. & Cuthbertson, A. M. *Nature* **318**, 361–363 (1985).
- Oeschger, H. *et al.* in *Climate Processes and Climate Sensitivity*, AGU Geophys. Monogr. Ser. Vol. 29 (eds Hansen, J. E. & Takahashi, T.) 299–306 (1984).
- Lorius, C., Jouzel, J., Raynaud, D., Hansen, J. & Le Treut, H. *Nature* **347**, 139–145 (1990).
- Weyl, P. *Met. Monogr.* **8**, 37–62 (1968).
- Rooth, C. *Prog. Oceanogr.* **11**, 131–149 (1982).
- Manabe, S. & Stouffer, R. J. *J. Clim.* **1**, 841–849 (1988).
- Broecker, W. S. & Denton, G. H. *Geochim. cosmochim. Acta* **53**, 2465–2501 (1989).
- Staffelbach, T., Stauffer, B., Sigg, A. & Oeschger, H. *Tellus* **43**, 91–96 (1991).
- Neftel, A., Oeschger, H., Staffelbach, T. & Stauffer, B. *Nature* **331**, 609–611 (1988).
- Boyle, E. A. & Keigwin, L. *Nature* **330**, 35–40 (1987).
- Jansen, E. & Veum, T. *Nature* **343**, 612–616 (1990).
- Oppo, D. W. & Fairbanks, R. G. *Earth planet. Sci. Lett.* **86**, 1–15 (1987).
- Curry, W. B., Duplessy, J. C., Labeyrie, L. D. & Shackleton, N. J. *Paleoceanography* **3**, 317–342 (1988).
- Belanger, P. E., Curry, W. B. & Matthews, R. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 205–220 (1981).
- Charles, C. D. & Fairbanks, R. G. in *The Geologic History of the Polar Oceans: Arctic vs Antarctic* (eds Bleil, U. & Thiede, J.) 519–538 (1990).

- Boyle, E. A. *Eos* **71**, 1357–1358 (1990).
- Boyle, E. A. & Rosenthal, Y. *Eos* **72**, 265 (1991).
- Bard, E. *Paleoceanography* **3**, 636–645 (1988).
- Labracherie, M. *et al.* *Paleoceanography* **4**, 629–638 (1989).
- Zahn, R., Pedersen, T. F., Bornhold, B. D. & Mix, A. C. *Paleoceanography* **6**, 543–560 (1991).
- Duplessy, J.-C. *et al.* *Paleoceanography* **3**, 343–360 (1988).
- Broecker, W. S. *J. geophys. Res.* **84**, 3218–3226 (1979).
- Wust, G. *Stratification and Circulation of the Antillean-Caribbean Basins* (Columbia University Press, New York, 1964).
- Ammann, B. & Lotter, A. F. *Boreas* **18**, 109–126.
- Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. *Nature* **345**, 405–410 (1990).
- Broecker, W. S., Bond, G., Klas, W., Bonani, G. & Wolff, W. *Paleoceanography* **5**, 469–478 (1990).
- Rind, D., Peteet, D., Broecker, W. S., McIntyre, A. & Ruddiman, W. *Clim. Dynam.* **1**, 3–33 (1986).
- Leventer, A., Williams, D. F. & Kennett, J. P. *Earth planet. Sci. Lett.* **59**, 11–17 (1982).
- Broecker, W. S. *et al.* *Radiocarbon* **32**, 119–133 (1990).
- Fairbanks, R. G., Charles, C. D. & Wright, J. W. in *Four Decades of Radiocarbon: An Interdisciplinary Approach* (eds Long, A. & Kra, R.) 473–500 (Springer, 1992).
- Van De Water, P. *Geol. Soc. Am. Abstr. Prog.* **352** (1991).
- Kroonnick, P. *Deep Sea Res.* **32**, 57–84 (1985).
- Fairbanks, R. G. *Paleoceanography* **6**, 937–948 (1990).

ACKNOWLEDGEMENTS. We thank T. M. Jull and D. Donohue at the Arizona Accelerator Facility for collaboration, and W. Broecker, R. Mortlock, D. Peteet, N. Slowey and J. Wright for discussions. This research was funded by NSF.



# Identification of features due to $\text{H}_3^+$ in the infrared spectrum of supernova 1987A

Steven Miller\*, Jonathan Tennyson\*, Stephen Lepp†  
& Alexander Dalgarno†

\* Department of Physics and Astronomy, University College London,  
Gower Street, London WC1E 6BT, UK

† Harvard-Smithsonian Center for Astrophysics, 60 Garden Street,  
Cambridge, Massachusetts 02138, USA

THE molecular ion  $\text{H}_3^+$  occupies a central position in theoretical models of interstellar chemistry<sup>1,2</sup>. It forms readily in hydrogen-rich interstellar gas clouds when ionized  $\text{H}_2$  reacts with neutral  $\text{H}_2$ . The  $\text{H}_3^+$  ion can then donate a proton to oxygen, carbon and other heavy atoms. From this beginning nearly 100 different interstellar molecules are formed, such as the reactive hydroxyl radical OH, neutral CO, ethanol, the linear polyacetylenes  $\text{HC}_x\text{N}$  and cyclic species such as cyclopropenylidene,  $\text{C}_3\text{H}_3$ . But in spite of a decade of searching<sup>3,4</sup>,  $\text{H}_3^+$  has eluded detection except in the hydrogen-rich atmosphere of Jupiter<sup>5</sup>. Here we present evidence for the presence of  $\text{H}_3^+$  in the envelope of supernova 1987A, where it is produced by an unusual chemistry in which excited hydrogen atoms are the main source of molecule formation. Infrared spectra of SN1987A (ref. 6) in the first few hundred days after the explosion contain two previously unidentified peaks which we attribute to  $\text{H}_3^+$ . Chemical modelling produces quantities of  $\text{H}_3^+$  consistent with the observed peak intensities, and also predicts significant amounts of  $\text{HeH}^+$ , which may be responsible for some weaker features in the spectra.

The infrared L-window spectra of supernova (SN) 1987A between 2.95 and 4.15  $\mu\text{m}$  are dominated by three hydrogen recombination lines<sup>6</sup>. But from day 110 onwards, there are strong, hitherto unidentified, features at 3.409 (3.41) and 3.533 (3.53)  $\mu\text{m}$ . These are precisely the wavelengths at which  $\text{H}_3^+$  announces its presence most strongly in Jupiter<sup>7,8</sup>. These features are clearest in the spectrum for day 192. Treating the mass of emitting  $\text{H}_3^+$ ,  $M_e$ , the velocity width at half maximum,  $V_w$ , and the temperature,  $T_{\text{exc}}$ , as disposable parameters, we have fitted the two features to emission from  $\text{H}_3^+$  using the observed frequencies and calculated transition probabilities<sup>9,10</sup> and an ortho-para ratio of 1.0. The background continuum was accommodated using a form similar to that of Meikle *et al.*<sup>6</sup> for the J, H and K windows:

$$B(\lambda) = a + b(\lambda - 3.60)^{-2.5} \quad (1)$$

where  $\lambda$  is the wavelength in micrometres. The fit produced values of  $M_e = 5.1 \times 10^{-8} M_\odot$ ,  $T_{\text{exc}} = 2,050 \text{ K}$ ,  $V_w = 3,200 \text{ km s}^{-1}$ , and  $a = 0.0$  and  $b = 0.6808$  in units of  $10^{-8} \text{ erg s}^{-1} \text{ cm}^{-2} \mu\text{m}^{-1}$  (Fig. 1).

We have propagated the spectrum of  $\text{H}_3^+$  obtained from our fitted values through the L window. For wavelengths longer than Pfund  $\delta$  at 3.318  $\mu\text{m}$ , the observed spectrum is entirely consistent with our fit, with features resulting from  $\text{H}_3^+$  contributing to the observed intensity in regions other than those specifically fitted. But at shorter wavelengths we predict a strong feature at 3.2  $\mu\text{m}$  and weaker peak centred on 3.02  $\mu\text{m}$ , which are not observed. These wavelengths correspond to regions of strong atmospheric water absorption, however, which may account for the discrepancy between the calculated and observed spectra.

But there may also be other contributors to the spectra that would alter the  $\text{H}_3^+$  parameters we have derived. In particular, Lucy *et al.*<sup>11</sup> have suggested that the 3.41- $\mu\text{m}$  feature results from the  $5^2\text{S}-4^2\text{P}^0$  doublet of sodium. This close doublet cannot, however, account for the width of the 3.41- $\mu\text{m}$  feature, and their interpretation encounters difficulties that can perhaps be

TABLE 1  $\text{H}_3^+$  and  $\text{HeH}^+$  mass

| Epoch (days) | $\text{H}_3^+ (M_\odot)$ | $\text{HeH}^+ (M_\odot)$ |
|--------------|--------------------------|--------------------------|
| 100          | $8.8 \times 10^{-8}$     | $1.4 \times 10^{-8}$     |
| 150          | $1.2 \times 10^{-7}$     | $2.2 \times 10^{-8}$     |
| 200          | $1.1 \times 10^{-7}$     | $2.5 \times 10^{-8}$     |
| 250          | $8.3 \times 10^{-8}$     | $2.5 \times 10^{-8}$     |
| 300          | $5.3 \times 10^{-8}$     | $2.2 \times 10^{-8}$     |
| 350          | $3.3 \times 10^{-8}$     | $1.8 \times 10^{-8}$     |

resolved by the presence of another emitter. We can accommodate a contribution from a second emitter, such as sodium, by reducing the  $\text{H}_3^+$  excitation temperature. In Fig. 2 we show the  $\text{H}_3^+$  spectrum for  $T_{\text{exc}} = 1,000 \text{ K}$ . Lowering  $T_{\text{exc}}$  has the additional advantage that the predicted feature at 3.2  $\mu\text{m}$  is much weaker and the discrepancy between the computed and observed spectra around 3.02  $\mu\text{m}$  is removed.

The ejecta of SN1987A consist of an envelope almost entirely composed of hydrogen and helium, and an inner core region in which the heavier elements are to be found. Densities are low in the supernova envelope, and local thermodynamic equilibrium is unlikely to prevail. Accurate values for the total mass of  $\text{H}_3^+$  will require a comprehensive study of the excitation conditions. But, given that the excitation temperature 1,000–2,000 K is similar to the kinetic temperature (which cannot exceed 3,000 K, as molecules are rapidly destroyed at temperatures higher than this), a total  $\text{H}_3^+$  mass of  $\sim 10^{-7} M_\odot$  is a reasonable estimate.

A mass of this order can be produced in the envelope of SN1987A by a chemistry initiated by the deposition of  $\gamma$ -rays. Recombination line profiles<sup>12,13</sup> show that hydrogen is not confined to the outer envelope of SN1987A but is mixed through the ejecta. In a pure hydrogen gas,  $\text{H}_3^+$  is formed by the reaction



The  $\text{H}_2^+$  ions are formed by radiative association, and the  $\text{H}_2$  molecules by charge transfer from H to  $\text{H}_2^+$  and by a negative ion sequence. In the supernova envelope, a large population

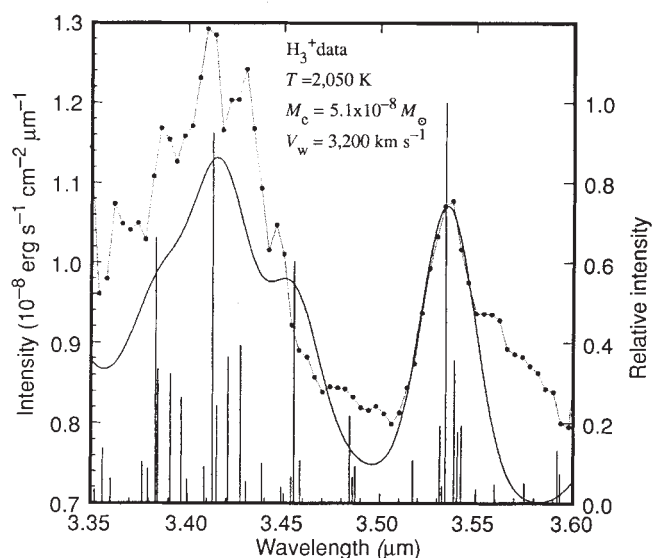
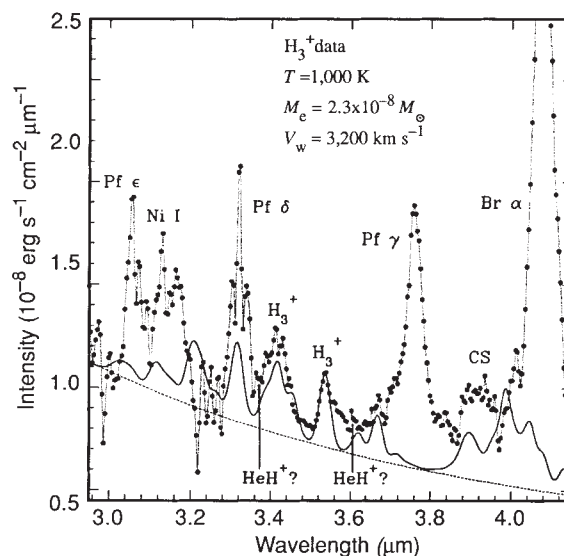
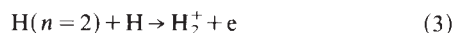


FIG. 1 Fit of the computed  $\text{H}_3^+ \nu_2 \rightarrow 0$  spectrum (solid curve) to the L-window spectrum of SN1987A on day 192 (ref. 6) (filled circles) from 3.35 to 3.60  $\mu\text{m}$ . Line positions and relative intensities of the individual  $\text{H}_3^+$  transitions that blend to form the 3.41- and 3.53- $\mu\text{m}$  features are shown. The 3.53- $\mu\text{m}$  feature results from a blend of two R(2) and two R(3) transitions in the  $\nu_2 \rightarrow 0$  band of  $\text{H}_3^+$  with individual line positions ranging from 3.531 to 3.543  $\mu\text{m}$ . The broader 3.41- $\mu\text{m}$  band is due to three separate blends, R(5), centred on 3.385  $\mu\text{m}$ , another R(3) blend at 3.420  $\mu\text{m}$  and R(4) at 3.455  $\mu\text{m}$ , comprising nine lines in all.

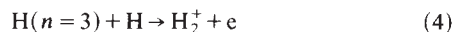
FIG. 2 Propagation of the fitted  $H_3^+$  spectrum (solid curve) through the L window data for day 192 (filled circles) assuming  $T_{exc}=1,000$  K. The dashed curve represents the continuum background. Assignments other than  $H_3^+$  and  $HeH^+$  are from ref. 6.



exists of hydrogen in the excited  $n=2$  and  $n=3$  states. These excited states contribute considerably to the formation of  $H_2^+$ . The process

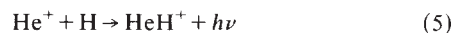


which has a threshold of 0.75 eV is a principal formation mechanism in hot stellar winds (J. M. C. Rawlings, J. E. Drew & M. J. Barlow, manuscript in preparation). The process



is exothermic and together with reaction (3) is an important source of molecules in the supernova.

The presence of helium enhances the formation of  $H_2^+$  and  $H_2$ , and through the reaction<sup>14</sup>



leads to a significant abundance of  $HeH^+$ .

In calculating the molecular abundances in the envelope gas, we have adopted the rate constants and physical conditions used by Xu *et al.*<sup>13</sup> in modelling the hydrogen recombination spectra. They assumed a temperature of 3,000 K for the gas. At this temperature associative ionization, reaction (3), is the chief source of  $H_2^+$  and ultimately  $H_3^+$ . The  $n=2$  level is populated by collisions from energetic electrons and by recombinations, and removed by two-photon decay, collisions with thermal electrons and ionization by continuum photons. Ionization from  $n=2$  is important in the ionization balance of the hydrogen gas.

The abundance of  $H_3^+$  depends critically on the rate coefficients of associative ionization, reactions (3) and (4), and of dissociative recombination



about which there is considerable uncertainty<sup>15</sup>. The calculated abundances are correspondingly uncertain. If we assume reactions (3), (4), (6) and (7) are all fast<sup>15</sup>, the chemical model yields

the abundances of  $H_3^+$  and  $HeH^+$  given in Table 1 for various times. The  $H_3^+$  mass at day 200 is in reasonable agreement with that deduced from the spectrum at day 192. Had we assumed that the dissociative recombination rates were slow<sup>16</sup>, however, the model would have produced an  $H_3^+$  mass around  $10^{-5} M_\odot$ , far more than the observed spectrum could accommodate.

Also shown in Table 1 is the mass of  $HeH^+$  as a function of time, showing that the model produces  $\sim 40$  times as much  $H_3^+$  as  $HeH^+$  by number. There exist in the spectra weak features which may correspond to the  $HeH^+$  vibrational transitions at 3.364  $\mu m$  and 3.607  $\mu m$ .

The L-window spectra of ref. 6 at days other than day 192 do not provide data of sufficient quality to enable us to carry out detailed spectroscopic fitting. In Table 2, however, we compare the intensities of the 3.53- $\mu m$  feature with those of H I Brackett  $\alpha$  and Pfund  $\gamma$  as measured by Meikle *et al.*<sup>6</sup>. The 3.53- $\mu m$  emission seems to track these recombination lines, as expected from the chemical scheme outlined above, lending further weight to our identification of  $H_3^+$ .

We conclude that our fit of  $H_3^+$  to the supernova data for day 192 produces a spectral profile consistent with that observed. It requires a mass of  $\sim 10^{-7} M_\odot$ , which can be produced by a plausible chemical model. The high abundance of  $H_3^+$  shows that minimal microscopic mixing of the heavy elements has taken place, as they would prevent the formation of  $H_3^+$  by destroying  $H_2$  in endothermic reactions. The existence of  $H_3^+$  places an upper limit of  $\sim 3,000$  K on the temperature of the envelope, a temperature which must be achieved in the absence of heavy elements as cooling agents. Although a sodium doublet may contribute to the 3.41- $\mu m$  emission, the features at 3.53, 3.36 and 3.61  $\mu m$  were unassigned previously. We believe that the observational and theoretical evidence strongly supports the presence of  $H_3^+$  in SN1987A and weakly supports the presence of  $HeH^+$ . If correct, we have now identified  $H_3^+$  beyond the Solar System and also identified extraterrestrial  $HeH^+$ .  $\square$

Received 17 July; accepted 8 November 1991.

1. Suzuki, H. *Prog. theor. Phys.* **62**, 936-947 (1979).
2. Herbst, E. & Klemperer, W. *Astrophys. J.* **185**, 505-513 (1973).
3. Geballe, T. R. & Oka, T. *Astrophys. J.* **342**, 855-859 (1989).
4. Black, J. H., van Dishoeck, E. F. & Woods, R. C. *Astrophys. J.* **358**, 459-467 (1990).
5. Drossart, P. *et al.* *Nature* **340**, 539-541 (1989).
6. Meikle, W. P. S., Allen, D. A., Spyromilio, J. & Varani, G.-F. *Mon. Not. R. astr. Soc.* **238**, 193-223 (1989).
7. Oka, T. & Geballe, T. R. *Astrophys. J.* **351**, L53-56 (1990).
8. Maillard, J.-P., Drossart, P., Watson, J. K. G., Kim, S. J. & Caldwell, J. *Astrophys. J.* **363**, L37-40 (1990).
9. Miller, S. & Tennyson, J. *Astrophys. J.* **335**, 486-490 (1988).
10. Kao, L., Oka, T., Miller, S. & Tennyson, J. *Astrophys. J. Suppl.* **77**, 317-329 (1991).
11. Lucy, L. B., Danziger, I. J. & Gouffes, C. *Astr. Astrophys.* **243**, 223-229 (1991).
12. Phillips, M. M. & Williams, R. E. in *Supernovae* (ed. Woosley, S. E.) 36 (Springer, New York, 1991).
13. Xu, Y., McCray, R., Oliva, E. & Randich, S. *Astrophys. J.* (in the press).

TABLE 2 Comparison of 3.53- $\mu m$  intensity with H I Br  $\alpha$  and Pf  $\gamma$

| Epoch (days) | 3.53 $\mu m$ | Br $\alpha$ | Pf $\gamma$ |
|--------------|--------------|-------------|-------------|
| 110          | 2.2          | 18.5        | 6.5         |
| 192          | 1.0          | 16.0        | 5.5         |
| 255          | 0.35         | 7.4         | 2.5         |
| 285          | 0.30         | —           | 2.4         |
| 349          | 0.17         | 5.9         | 1.0         |

Units:  $10^{-10} \text{ erg s}^{-1} \text{ cm}^{-2}$ .



14. Lepp, S., Dalgarno, A. & McCray, R. *Astrophys. J.* **358**, 262–267 (1990).  
 15. Amano, T. *Astrophys. J.* **329**, L121–L124 (1988).  
 16. Adams, N. G. & Smith, D. in *Astrochemistry* (ed. Vardya, M. S. & Tarafdar, S. P.) 1 (Reidel, Dordrecht, 1986).

ACKNOWLEDGEMENTS. We acknowledge the assistance of P. Meikle in the preparation of this paper, and thank him and R. Cummings for the latest processed data. We also thank D. Allen and H. Schild for discussions. J.T. and S.M. acknowledge support from the SERC. A.D. and S.L. acknowledge support from the NSF, Division of Astronomical Sciences.

## Observation of gradual brightening of P Cygni due to stellar evolution

Mart J. H. de Groot\* & Henny J. G. L. M. Lamers†

\* Armagh Observatory, College Hill, Armagh BT61 9DG, Northern Ireland

† Astronomical Institute and SRON Laboratory for Space Research, Sorbonnelaan2, 3584 CA, Utrecht, The Netherlands

WITH the exception of supernova outbursts, or changes in the pulsational periods of Cepheid variables arising from their changing internal structure, the evolution of stars is generally much too slow to have been detected during the era of modern astronomy. A few stars, such as the red supergiants  $\rho$  Cas and RW Cep, have shown spectral changes on a timescale of decades which have been attributed to evolutionary effects<sup>1</sup>, but these changes are mostly erratic and the evolutionary interpretation is uncertain. Here we report an analysis of modern and historical (back to about AD 1700) photometric measurements of the star P Cygni. We find a steady change of apparent brightness, and argue that it is due to evolution of the star. The change is about twice as fast as standard models predict, but the difference may be due to mis-estimation of the star's mass or inadequate treatment of atmospheric expansion in the stellar models.

P Cygni (HD193237) is a luminous blue variable of spectral type B1 Ia+ with a present mean visual brightness of 4.8 mag. Its effective temperature is  $T_{\text{eff}} = 19,300 \pm 700$  K, and its luminosity,  $L$ , is given by  $\log L/L_{\odot} = 5.86 \pm 0.10$  (ref. 2). A comparison of these values with predicted evolutionary tracks<sup>3</sup> suggests an initial mass of  $50 \pm 10 M_{\odot}$  and a present mass of  $30 \pm 10 M_{\odot}$ .

P Cyg's photometric history has been recorded in several bibliographies<sup>4–6</sup>. Mueller and Hartwig's list is a careful interpretation of the many historical observations of P Cyg with the stellar magnitude estimates reduced to the system of the Potsdam Durchmusterung. We have transformed these photometric magnitudes into the V-magnitude scale to couple the historical and modern observations into one system.

Measurements of the visual magnitudes of P Cyg between 1700 and 1990 are listed in Table 1. The eighteenth- and nineteenth-century magnitudes of P Cyg are based on optical estimates through small telescopes by many observers. The accuracy of the individual observations is estimated to be  $\sim 0.3$  mag. Apart from this observational uncertainty, there is the irregular intrinsic brightness variation of P Cyg. This is assumed to be the same as the modern brightness variations in V of 0.2 mag (ref. 7). We have averaged the magnitude estimates of the eighteenth- and nineteenth-century over periods of between 8 and 29 years, depending on the frequency of the observations at each epoch. This has the advantage that it produces a long-term light curve essentially free from the rapid variations, and it eliminates as far as possible the small differences between individual observers.

The observations between 1898 and 1917 contain the first photoelectric photometry of P Cyg, with the magnitudes transformed to the standard system by Mueller and Hartwig<sup>4</sup>. Between 1920 and 1950, observers used a variety of photoelectric detectors and almost any combination of filters, until the international UBV system was established in 1953. Therefore we have not tried to reduce observations in this period to the scale of

the V-magnitude. The mean value of the unspecified number of photometric observations of P Cyg between 1952 and 1954 by Hiltner<sup>8</sup> in the UBV-systems is listed under epoch 1953. The magnitude at epoch 1966 is the mean of only two observations of P Cygni in 1965 and 1966 with the very accurate 13-colour system of Johnson and Mitchell<sup>9</sup>. The magnitude in their (52)-filter (centred at 520 nm) was converted to the V-magnitude by means of the transformation formula of ref. 10. Percy and Welch<sup>11</sup> observed the star on 11 nights in 1982. The visual magnitude at epoch 1988 is derived from the photoelectric measurements by members of the American Association of Variable Star Observers between 1985 and 1990 (ref. 12).

The mean visual magnitudes are plotted in Fig. 1, which clearly shows the trend of the increasing visual brightness. The weighted least-square fit through the data is

$$V = 5.24 - (1.54 \pm 0.20) \times 10^{-3}(t - 1700) \quad (1)$$

where  $t$  is the date in years, which corresponds to a brightening of  $dV/dt = -0.15 \pm 0.02$  mag per century.

The theory of evolution of massive stars predicts that the evolution occurs at almost constant  $L$  throughout most of the star's life. This is because massive stars do not have degenerate cores and because the energy transport is mainly by radiation, except in the red-supergiant phase. A constant luminosity implies that the changes in  $V$  are due to the redistribution of the emitted radiation over the spectrum: a change in  $V$  reflects the change in the star's bolometric correction (BC) and  $T_{\text{eff}}$ . A change in  $T_{\text{eff}}$  is related to a change in  $V$  according to

$$\frac{d \log(T_{\text{eff}})}{dt} = \frac{d \log(T_{\text{eff}})}{d(\text{BC})} \frac{d(\text{BC})}{dt} = -0.027 \pm 0.004 \quad (2)$$

in mag per century. We adopted the empirical relation between BC and  $T_{\text{eff}}$  for supergiants of luminosity class Ia, tabulated by Schmidt-Kaler<sup>13</sup>, which gives  $d(\text{BC})/d \log(T_{\text{eff}}) = -5.63$  mag near  $T_{\text{eff}} \approx 20,000$  K. The observed gradual brightening of P Cyg thus implies a steady decrease in  $T_{\text{eff}}$  of  $6 \pm 1\%$  per century. The decrease in  $T_{\text{eff}}$  is related to an increase in radius because  $L \sim R_*^2 T_{\text{eff}}^4$  is constant. The timescale  $\tau(R_*)$  in which  $R_*$  increases by a factor  $e$  is  $320 \pm 60$  yr.

The location of P Cyg in the Hertzsprung–Russell diagram (HRD) is shown in Fig. 2, which also shows some predicted evolutionary tracks of massive stars with mass loss and convective overshooting<sup>3</sup>. P Cyg lies in a region of the HRD that is crossed twice during the evolution of a massive star. The first crossing occurs immediately after the main-sequence phase when the energy of the star is generated by a hydrogen-burning shell and by contraction of the core. The outer layers react by

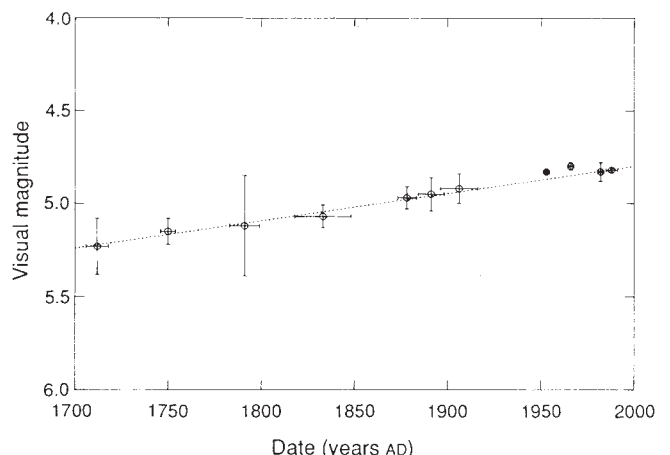


FIG. 1 Visual magnitude of P Cyg as a function of time. Dotted line, least-square fit given by equation (1) which shows the gradual photometric brightening of  $0.15 \pm 0.02$  mag per century.

TABLE 1 P Cygni's mean light curve from 1700 to 1990

| Date (yr) | Epoch (yr) | Mag (V)         | Number of observations |
|-----------|------------|-----------------|------------------------|
| 1704-1715 | 1712       | $5.23 \pm 0.15$ | 4                      |
| 1746-1754 | 1750       | $5.15 \pm 0.07$ | 2                      |
| 1779-1795 | 1791       | $5.12 \pm 0.27$ | 8                      |
| 1814-1843 | 1833       | $5.07 \pm 0.06$ | 3                      |
| 1872-1882 | 1878       | $4.97 \pm 0.06$ | 3                      |
| 1884-1897 | 1891       | $4.95 \pm 0.09$ | 7                      |
| 1898-1917 | 1906       | $4.92 \pm 0.08$ | 3                      |
| 1952-1954 | 1953       | $4.83 \pm 0.01$ | ?                      |
| 1965-1966 | 1966       | 4.80            | 2                      |
| 1982      | 1982       | $4.83 \pm 0.05$ | 11                     |
| 1985-1990 | 1988       | $4.82 \pm 0.01$ | 476                    |

expanding so that the star moves to the right in the HRD at constant luminosity. The second crossing occurs in a later phase when the star has lost so much mass that it can no longer sustain an extended convective envelope above the H-burning shell. In that phase the envelope contracts and the star moves to the left of the HRD to become a Wolf-Rayet star.

The observed decrease of  $T_{\text{eff}}$  shows that P Cyg is in the expansion phase after the main sequence. This expansion is expected to occur on the Kelvin-Helmholtz timescale for contraction

$$\tau_{\text{KH}} = GM^2/R_*L \approx 520 \text{ yr} \quad (3)$$

which is of the same order of magnitude as the observed timescale of the increase of  $R_*$ , of  $\tau(R_*) = 320 \text{ yr}$ . This supports the suggestion that the star is in the H-shell burning phase.

The predicted evolutionary speed of a star with the luminosity of P Cygni in the H-shell burning phase near  $T_{\text{eff}} = 20,000 \text{ K}$  is  $d \log(T_{\text{eff}})/dt = -0.013 \text{ per century}^3$ , which implies a decrease in  $T_{\text{eff}}$  of only 3% per century. This is about half the observed value of 6% per century. So the observed evolution speed of P Cyg is twice as fast as predicted. The discrepancy between the observed and predicted evolutionary speed might be due to two effects: (1) The calculated evolutionary speed might be wrong because the outer layers of the star expand faster than predicted. In fact, recent calculations of the evolution of luminous stars by A. Maeder (personal communication) suggest that dynamical effects due to radiation pressure in the envelope of the star result in a more rapid expansion of the star and consequently also in a more rapid decrease in  $T_{\text{eff}}$  than predicted

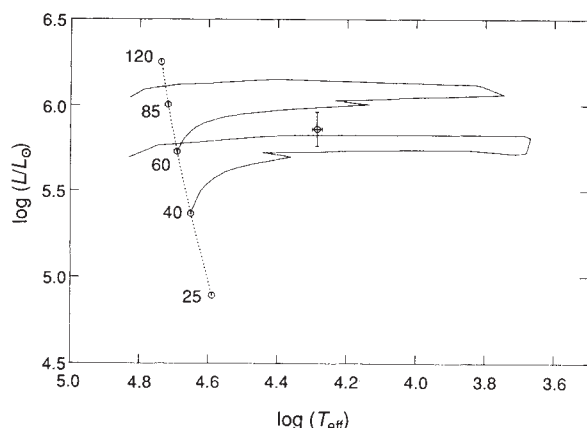


FIG. 2 Location of P Cyg in the Hertzsprung-Russell diagram compared with the predicted evolutionary tracks<sup>15</sup> of two stars with initial masses of 40 and 60  $M_{\odot}$ . The zero-age main sequence (dotted) for stars of various initial masses is also shown. The observed visual brightening of P Cyg indicates that it is evolving to the right in this diagram at a rate twice as large as predicted by models of stellar evolution.

by the quasi-hydrostatic evolution calculations<sup>3</sup>. (2) The mass of the star, derived from its luminosity, might be overestimated. In fact, Pauldrach and Puls<sup>14</sup> derived a surface gravity for P Cyg of  $\log g = 2.04 \pm 0.01$  from detailed modelling of its stellar wind. Assuming a luminosity of  $\log L/L_{\odot} = 5.86 \pm 0.1$ , they derived a mass of  $M = 23 \pm 2 M_{\odot}$ . This is about a factor of 0.6 smaller than the mass of  $40 \pm 4 M_{\odot}$  derived from the predicted evolution for a star of P Cyg's  $T_{\text{eff}}$  and  $L$  if it is evolving to the right in the HRD<sup>15</sup>. If the core mass is also smaller than predicted, by the same factor of 0.6, the timescale for the core contraction and the envelope expansion will be about a factor two shorter than predicted by the evolutionary calculations of Maeder and Meynet<sup>3</sup>, as this timescale is proportional to  $M_{\text{core}}^2$ . This would bring the predicted timescale for the expansion into agreement with the observed brightening of P Cyg.  $\square$

Received 21 October; accepted 15 November 1991.

- de Jager, C. *The Brightest Stars*, 100 (Reidel, Dordrecht, 1980).
- Lamers, H. J. G. L. M., de Groot, M. J. H. & Cassatella, A. *Astr. Astrophys.* **128**, 299-310 (1983).
- Maeder, A. & Meynet, G. *Astr. Astrophys.* **182**, 243-263 (1987).
- Mueller, G. & Hartwig, E. in *Geschichte und Literatur des Lichtwechsels der Veränderlichen Sterne*, 444 (Universitätssternwarte Berlin-Babelsberg, 1918).
- Prager, R. in *Geschichte und Literatur des Lichtwechsels der Veränderlichen Sterne*, 61 (Universitätssternwarte Berlin-Babelsberg, 1936).
- Schneller, R. in *Geschichte und Literatur des Lichtwechsels der Veränderlichen Sterne*, 258 (Akademie, Berlin, 1957).
- De Groot, M. J. H. in *Physics of Luminous Blue Variables* (eds Davidson, K. et al.) 257-258 (Kluwer, Dordrecht, 1989).
- Hiltner, W. A. *Astrophys. J. suppl. Ser.* **2**, 389-399 (1956).
- Johnson, H. L. & Mitchell, R. I. *Rev. Mex. Astr. Astrofis.* **1**, 299-324 (1975).
- Mitchell, R. I. & Johnson, H. L. *Comm. lunar planet. Lab.* **8**, 1-49 (1970).
- Percy, J. R. & Welch, D. L. *Publ. astr. Soc. Pacific* **95**, 491-505 (1983).
- Percy, J. R. et al. *Astr. Astrophys.* **191**, 248-252 (1988).
- Schmidt-Kaler, Th. in *Landolt-Bornstein, New Series, Group VI, Vol. 2b*, 452 (Springer, Berlin, 1982).
- Pauldrach, A. W. A. & Puls, J. *Astr. Astrophys.* **237**, 409-424 (1990).
- Maeder, A. *Astr. Astrophys. Suppl.* **84**, 139-177 (1990).

## Territory covered by $N$ diffusing particles

Hernan Larralde\*, Paul Trunfio\*, Shlomo Havlin\*†, H. Eugene Stanley\* & George H. Weiss†

\* Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

† Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892, USA

THE number of distinct sites visited by a random walker after  $t$  steps is of great interest<sup>1-21</sup>, as it provides a direct measure of the territory covered by a diffusing particle. Thus, this quantity appears in the description of many phenomena of interest in ecology<sup>13-16</sup>, metallurgy<sup>5-7</sup>, chemistry<sup>17,18</sup> and physics<sup>19-22</sup>. Previous analyses have been limited to the number of distinct sites visited by a single random walker<sup>19-22</sup>, but the (nontrivial) generalization to the number of distinct sites visited by  $N$  walkers is particularly relevant to a range of problems—for example, the classic problem in mathematical ecology of defining the territory covered by  $N$  members of a given species<sup>13-16</sup>. Here we present an analytical solution to the problem of calculating  $S_N(t)$ , the mean number of distinct sites visited by  $N$  random walkers on a  $d$ -dimensional lattice, for  $d = 1, 2, 3$  in the limit of large  $N$ . We confirm the analytical arguments by Monte Carlo and exact enumeration methods. We find that there are three distinct time regimes, and we determine  $S_N(t)$  in each regime. Moreover, we also find a remarkable transition, for dimensions  $\geq 2$ , in the geometry of the set of visited sites. This set initially grows as a disk with a relatively smooth surface until it reaches a certain size, after which the surface becomes increasingly rough.

The formalism used to obtain the analytical results, and on which the exact enumeration<sup>20</sup> calculations are based, is as follows. We denote the probability that a single random walker



FIG. 1 Schematic illustration of the results for the number of distinct sites visited by  $N$  random walkers initially at the origin, indicating the fact that at different times  $t$  the behaviour is very different. Here,  $S_1$  is given by equation (7).

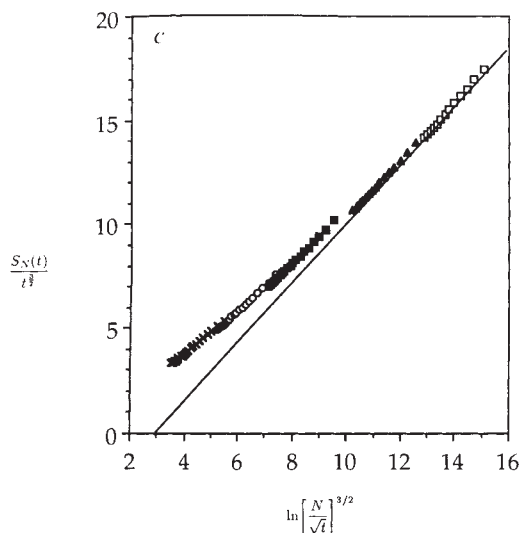
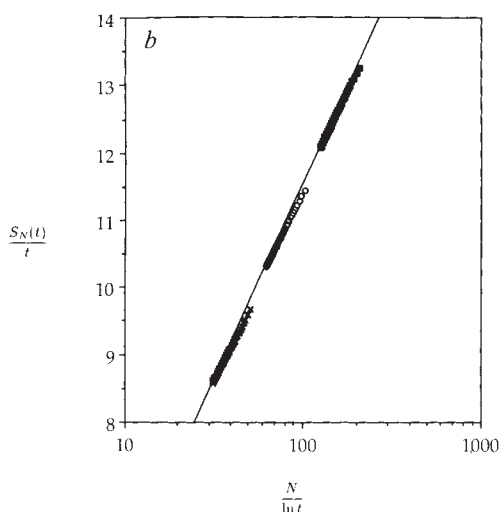
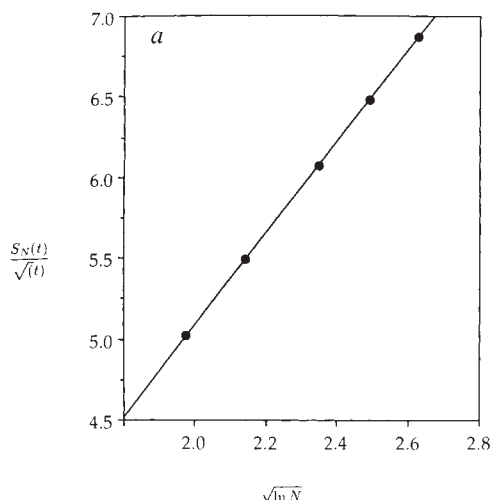
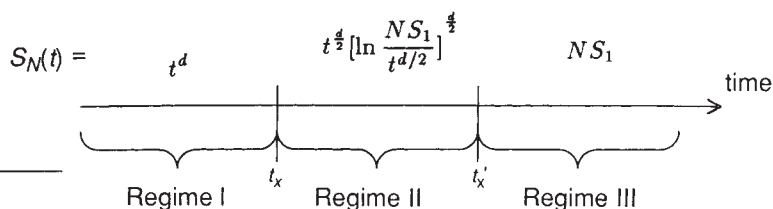


FIG. 2 Scaling plot demonstrating that equation (4) holds for (a) a one-dimensional system ( $d=1$ ), (b) a two-dimensional system ( $d=2$ ) and (c) a three-dimensional system ( $d=3$ ). Each data point in a and each of the different symbols in b and c correspond to different values of  $N$ .

is at site  $\mathbf{r}$  for the first time at step  $t$  by  $f_t(\mathbf{r})$ , and the probability that site  $\mathbf{r}$  has not been visited by this walker by step  $t$  as  $\Gamma_t(\mathbf{r})$ . Then

$$\Gamma_t(\mathbf{r}) = 1 - \sum_{t'=0}^t f_{t'}(\mathbf{r}) \quad (1)$$

Because the  $N$  walkers are independent, the probability that site  $\mathbf{r}$  has not been visited by any of the  $N$  random walkers is  $[\Gamma_t(\mathbf{r})]^N$ . The probability that site  $\mathbf{r}$  has been visited by at least one walker by step  $t$  is  $1 - [\Gamma_t(\mathbf{r})]^N$ . Thus the expected number of distinct sites visited by any of the  $N$  random walkers by step  $t$  is obtained by summing over all sites  $\mathbf{r}$  in the lattice

$$S_N(t) = \sum_{\mathbf{r}} \{1 - [\Gamma_t(\mathbf{r})]^N\} \quad (2)$$

We can now analyse the very short time behaviour of  $S_N(t)$  (regime I; see Fig. 1). As  $N$  tends to infinity, any number raised to the power  $N$  approaches zero if the number is less than one; hence  $[\Gamma_t(\mathbf{r})]^N$  tends to zero if it is possible that a walker may arrive at site  $\mathbf{r}$  by step  $t$ . From equation (2) we see that  $S_N(t)$  consists of all the sites that have non-zero probability of being visited by step  $t$ , and is therefore independent of  $N$ . As the walker can take only nearest-neighbour steps, it follows that for large  $N$ ,  $S_N(t)$  is given by the simple expression

$$S_N(t) \sim At^d \quad [t < t_x] \quad [\text{regime I}] \quad (3a)$$

where  $A$  depends on the lattice. This behaviour should hold so long as the number of accessible sites is much smaller than  $N$  or, more precisely, so long as  $1/P_{\min}(t) \ll N$ , where  $P_{\min}$  is the smallest non-zero occupation probability on the lattice at time  $t$ . Then  $P_{\min}(t) = z^{-t}$ , where  $z$  is the number of nearest neighbours of a site, so regime I must end at a crossover time  $t_x$  which scales logarithmically with  $N$ ,

$$t_x \sim \ln N \quad (3b)$$

Regime I is not present for small  $N$ , because it arises from the condition that all the accessible sites must be occupied by many walkers. For large  $N$ , the initial growth of the covered territory is very fast, yet the initial growth rate is independent of  $N$  (only the duration time  $t_x$  of this regime depends on  $N$ ).

To discuss times greater than  $t_x$ , we calculate  $S_N(t)$  from equation (2) using generating function techniques<sup>23-25</sup>. As we are now interested in the long-time behaviour, we use the continuum approximation to evaluate  $\Gamma_t(\mathbf{r})$ . This analysis leads to a compact scaling expression for  $S_N(t)$

$$S_N(t) \sim t^{d/2} f(x) \quad [t \gg t_x] \quad (4a)$$

Throughout this paper, the tilde ( $\sim$ ) denotes the fact that an equation holds for  $N$  and  $t$  both large. The scaled variable  $x$  is given by

$$x = \begin{cases} N & [d=1] \\ N/\ln t & [d=2] \\ N/\sqrt{t} & [d=3] \end{cases} \quad (4b)$$

and the scaling function  $f(x)$  by

$$f(x) = \begin{cases} (\ln x)^{d/2} & t_x \ll t \ll t'_x \quad [\text{regime II}] \\ x & t \gg t'_x \quad [\text{regime III}] \end{cases} \quad (5)$$

Again, these regimes are shown in Fig. 1. Here the second

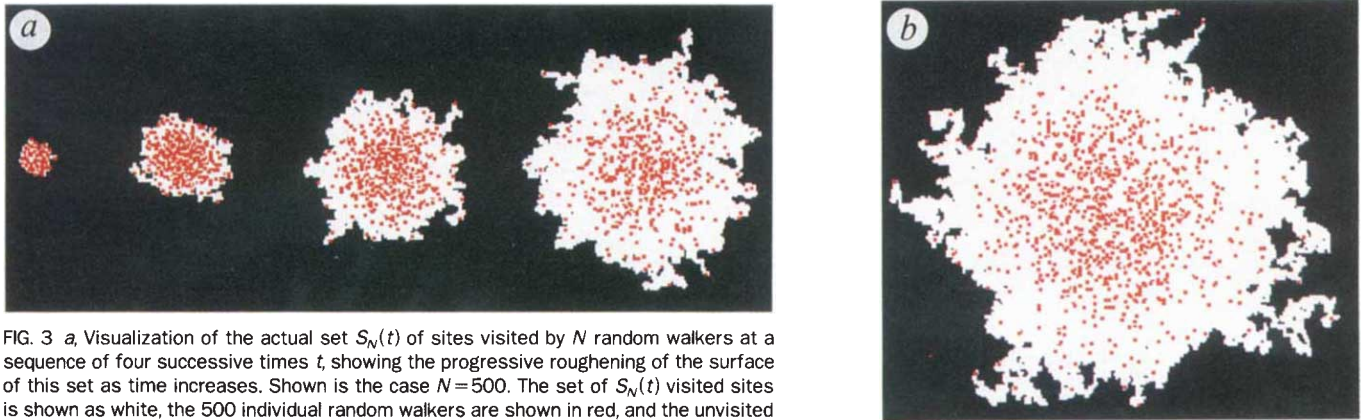


FIG. 3 *a*, Visualization of the actual set  $S_N(t)$  of sites visited by  $N$  random walkers at a sequence of four successive times  $t$ , showing the progressive roughening of the surface of this set as time increases. Shown is the case  $N=500$ . The set of  $S_N(t)$  visited sites is shown as white, the 500 individual random walkers are shown in red, and the unvisited 'virgin territory' is shown in black. *b*, The case  $N=1,000$  at late time, which demonstrates the part played by only a few individual random walkers in causing the roughening of the interface of the set  $S_N(t)$ .

crossover time  $t'_x$  is

$$t'_x \sim \begin{cases} \infty & [d=1] \\ e^N & [d=2] \\ N^2 & [d=3] \end{cases} \quad (6)$$

The predicted scaling form of equation (4) was tested by extensive calculations using the method of exact enumeration; results are shown in Fig. 2. Note that exact enumerations cannot readily be carried out to times long enough to see regime III because equation (6) implies that  $t'_x$  occurs at fairly large times (especially for  $d=2$ ).

The crossover from regime II to regime III can be understood in the following way. In regime II, the walkers are contained within a  $d$ -dimensional sphere of radius  $\xi \sim t^{1/2}$  (except for rare fluctuations). Hence  $S_N(t)$  must be bounded from above by the volume of this sphere,  $V(t) \sim t^{d/2}$ . A second upper bound on  $S_N(t)$  is  $NS_1(t)$ , where

$$S_1(t) \sim \begin{cases} t^{1/2} & [d=1] \\ t/\ln t & [d=2] \\ t & [d=3] \end{cases} \quad (7)$$

is the number of distinct sites visited by one random walker. A crossover in  $S_N(t)$  will occur if the system passes from one constraint to the other. For  $d=1$ ,  $V(t) < NS_1(t)$  for all  $t$ , so no crossover occurs, and regime II holds for arbitrarily large  $t$ , confirming the result of equation (6a) above. For  $d=2$  and 3, we find  $V(t) < NS_1(t)$  initially, but for sufficiently large  $t$ ,  $V(t) > NS_1(t)$ . Thus  $t'_x$  is obtained from the condition

$$V(t'_x) \sim NS_1(t'_x) \quad (8)$$

For  $d=2$ , equations (7) and (8) lead to  $t'_x \sim Nt'_x/\ln t'_x$ , so that  $t'_x \sim e^N$ ; this confirms the result (6b) above. Similarly, for  $d=3$ ,  $(t'_x)^{3/2} \sim Nt'_x$  implies  $t'_x \sim N^2$ , confirming the result (6c). Actually, following the same kind of reasoning, we can conclude that the crossover time to the final regime in any dimension higher than 2 will be given by

$$t'_x \sim N^{2/(d-2)} \quad [d > 2] \quad (9)$$

This result is a consequence of the fact that  $S_1(t) \sim t$  for any dimension larger than 2. One can interpret  $t'_x$  as the time up to which the walkers visit the same places very frequently. For times longer than  $t'_x$ , the walkers 'almost' do not see each other, and can be treated independently. Equation (9) shows the effect of the space dimension on  $t'_x$ ; the higher the dimension, the shorter the times at which walkers become 'independent'.

From Monte Carlo simulations for  $d=2$ , we find a remarkable transition in the geometry of the set of visited sites. This set initially grows in the shape of a disk with a relatively smooth surface; this occurs at very short times, during regime I. The

surface then becomes increasingly rough in regime II (see Fig. 3) as the walkers move further away from one another, until they are so sparse that they can be treated independently; this occurs in regime III. This phenomenon can be linked to  $S_N(t)$  by noting that the walkers can find new sites only on the surface of the set of visited sites. Accordingly, the 'growth rate'  $\partial S_N(t)/\partial t$  will be proportional to the number of surface sites. For  $d=2$  and 3, the number of surface sites increases rapidly as the surface roughens, thus leading to an increase in the growth rate of  $S_N(t)$  until it reaches its maximum possible value in regime III. For  $d=1$ , the surface always consists of two points (the two ends of the set of visited sites), so there is no roughening contribution to the growth rate of  $S_N(t)$ , which again agrees with the absence of regime III in the one-dimensional case.

The roughening of the surface of the set of visited sites (Fig. 3) can be understood as follows. We denote by  $P(r, t)$  the probability that a walker initially at the origin is found at a distance  $r$  from the origin at time  $t$ . Then the average number of walkers at position  $r$  at time  $t$  will be given by  $NP(r, t)$ . In regime I, we have  $NP(r, t) > NP_{\min}(t) \gg 1$ , so on average all accessible sites are multiply occupied. This set is roughly a sphere of radius  $t$ . A different situation occurs in regime II, in which there exists a distance  $\rho$  for which  $NP(\rho, t) = 1$ ; for  $r > \rho$ , the walkers are increasingly sparse. This results in the roughening of the surface of the set of visited sites. Thus regime I is characterized by filling a region of radius  $r_1 \sim t$ , whereas in regime II there are insufficient walkers to fill more than a region of radius  $r_2 \sim \sqrt{t}$ . The region between  $r_2 \sim \sqrt{t}$  and  $r_1 \sim t$  is characterized by walkers who become increasingly isolated from one another with time, leading to an increasingly rough exterior surface. This phenomenon may have been observed by Skellam<sup>13</sup>, who plotted contours delineating the advance of the muskrat population, and noted that the contours were initially smooth but later became rough (see Fig. 1 of ref. 13).

Finally, we note that the existence of the different time regimes is a consequence of the initial localization of all  $N$  walkers at the origin, which is not an unreasonable assumption in many cases of interest. Suppose, on the other hand, that walkers are initially localized in a box of linear size  $l$ . Then we can distinguish three types of behaviour for  $S_N(t)$ . If  $l \ll \xi(t_x) \sim \sqrt{\ln(N)}$ , we expect to observe all three growth regimes of  $S_N(t)$ . If  $\xi(t_x) \ll l \ll \xi(t'_x)$ , then the system would initially be in regime II, and only the crossover to regime III would occur. If  $l \gg \xi(t'_x)$ , then regime III will hold for all times.  $\square$

Received 9 October; accepted 5 November 1991.

1. Montroll, E. W. & West, B. J. in *Fluctuation Phenomena* (eds Montroll, E. W. & Lebowitz, J. L.) (North-Holland Personal Library, 1987).
2. Montroll, E. W. & Shlesinger, M. F. in *Nonequilibrium Phenomena II. From Stochastics to Hydrodynamics* (eds Lebowitz, J. L. & Montroll, E. W.) 1-121 (North-Holland, Amsterdam, 1984).



3. Dvoretzky, A. & Erdős, P. in *Proc. 2nd Berkeley Symp.* 33 (University of California, Berkeley, 1951).
4. Vineyard, G. H. *J. math. Phys.* **4**, 1191–1193 (1963).
5. Beeler, R. J. & Delaney, J. A. *Phys. Rev.* **A130**, 962–966 (1963).
6. Beeler, R. J. *Phys. Rev.* **A134**, 1396–1401 (1964).
7. Rosenstock, H. B. *Phys. Rev.* **A187**, 1166–1168 (1969).
8. Montroll, E. W. in *Stochastic Processes in Applied Mathematics XVI*, 193–220 (American Mathematical Society, Providence, 1964).
9. Montroll, E. W. & Weiss, G. H. *J. math. Phys.* **6**, 167–177 (1965).
10. Jain, N. C. & Orey, S. *Israel J. Math.* **6**, 373–380 (1968).
11. Henyey, F. S. & Seshadri, V. *J. chem. Phys.* **76**, 5530–5534 (1982).
12. Torney, D. C. *J. stat. Phys.* **44**, 49–66 (1986).
13. Skellam, J. G. *Biometrika* **38**, 196–218 (1951).
14. Skellam, J. G. *Biometrika* **39**, 346–362 (1952).
15. Pielou, E. C. *An Introduction to Mathematical Ecology* (Wiley-Interscience, New York, 1969).
16. Edelstein-Keshet, L. *Mathematical Models in Biology* (Random House, New York, 1988).
17. Smoluchowski, M. v. *Z. phys. Chem.* **29**, 129 (1917).
18. Rice, S. A. *Diffusion-Controlled Reactions* (Elsevier, Amsterdam, 1985).
19. Haus, J. W. & Kehr, K. W. *Physics Rep.* **150**, 263–416 (1987).
20. Havlin, S. & Ben-Avraham, D. *Adv. Phys.* **36**, 695–798 (1987).
21. Bouchaud, J.-P. & Georges, A. *Physics Rep.* **195**, 127–293 (1990).
22. Bunde, A. & Havlin, S. (eds) *Fractals and Disordered Systems* (Springer, Berlin, 1991).
23. Barber, M. N. & Ninham, B. W. *Random and Restricted Walks* (Gordon & Breach, New York, 1970).
24. Berg, H. C. *Random Walks in Biology* (Princeton University Press, 1983).
25. Weiss, G. H. & Rubin, R. J. *Adv. chem. Phys.* **52**, 363–475 (1983).

ACKNOWLEDGEMENTS. We thank M. Araujo, A. L. Barabasi, G. Huber, J. Lee and P. H. Poole for discussions, and CONACYT, ONR, NSF and the US-Israel Binational Foundation for support.

## Crumpled and collapsed conformations in graphite oxide membranes

Xin Wen\*, Carl W. Garland\*, Terence Hwa†, Mehran Kardar\*, Etsuo Kokufuta‡, Yong Li\*, Michal Orkisz\* & Toyochi Tanaka\*

\* Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Physics Department Harvard University, Cambridge, Massachusetts 02138, USA

‡ Institute of Applied Biochemistry, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

**MEMBRANES composed of bilayers of amphiphiles such as phospholipids generally exhibit two-dimensional liquid-like structure within the layers. When the constituent molecules of such a membrane are permanently cross-linked to each other, the membrane becomes less flexible, forming a two-dimensional solid. Solid membranes are expected to exhibit very different behaviour from their liquid counterparts<sup>1–3</sup>, including transitions between a two-dimensional flat phase, a crumpled phase of fractal dimension 2.5 and a compact, three-dimensional phase. Experimental evidence for the crumpled phase has, however, been lacking. As this phase was not observed in computer simulations<sup>4–6</sup>, it has been suggested that it may always be absent for self-avoiding (and therefore all real) membranes<sup>4–6</sup>. To the contrary, we report here the experimental observation of the crumpled conformation in an aqueous suspension of graphite oxide membranes. Static light scattering measurements indicate the presence of membrane conformations with a fractal dimension of  $2.54 \pm 0.05$ . As the intra-membrane affinity is enhanced by changing the composition of the solvent, the membranes collapse to a compact configuration.**

A natural state for a solid membrane is a flat configuration. Unlike in a liquid membrane which lacks internal shear elasticity, out-of-plane thermal undulations in a solid membrane are not energetically favoured, because they are accompanied by in-plane strains<sup>2</sup>. When the in-plane elasticity is sufficiently weak or the temperature high enough, it has been argued that a crumpled conformation with random orientations should form because the entropy gain from the increased number of bent and folded configurations more than compensates the internal energy cost<sup>7,8</sup>. Like the coiled polymer, the crumpled membrane is a tenuous fractal<sup>9</sup> with a dimension determined by the competition between the configurational entropy and steric constraints. If the affinity between the molecules is increased, the membrane

may collapse completely into a compact configuration<sup>10,11</sup>. Figure 1 provides a schematic representation of these three phases.

In computer simulations<sup>4–6</sup>, flat and compact<sup>10,11</sup> conformations have been realized by tuning simple potentials, but the intermediate crumpled phase has not been observed in equilibrium. This has led to the conjecture that a crumpled phase is always absent in self-avoiding membranes<sup>7</sup>. Experimental studies of solid membranes are therefore most desirable. Candidates for solid membranes include sheets of graphite oxide (GO)<sup>12–15</sup> and MoS<sub>2</sub> (ref. 16), the spectrin network extracted from red blood cells (C. Schmitt, personal communication) and polymerized liquid membranes<sup>17</sup>. Previous electron microscope images show that both GO<sup>15</sup> and MoS<sub>2</sub> (ref. 16) form irregular and folded configurations when dried on the microscope stage. Here we study conformations of GO membranes suspended in dilute solution.

GO membranes are synthesized by exfoliating bulk graphite with strong oxidizing agents<sup>12,13,15</sup>; we used potassium permanganate. The product of the oxidation reaction is GO, brownish yellow in colour, which consists of carbon layers bonded by oxygens<sup>14</sup>. Electron microscopy has shown that the GO membranes have a crystalline internal order<sup>15</sup> and can be as thin as one carbon layer<sup>14</sup>. Hydroxyl (OH) groups are bound to the surfaces of the GO membranes<sup>14</sup>. These hydroxyl groups, capable of forming hydrogen bonds with water molecules, are responsible for the compatibility of GO with aqueous solutions.

A light scattering study of the structure of membranes in solution requires some control over their sizes. This was achieved by using polycarbonate filters (Nucleopore). The GO suspension was sequentially passed through filters with 8- $\mu$ m and 3- $\mu$ m pore size. The membranes retained on the 3- $\mu$ m filter were redispersed in a solution, with the original buffer conditions. The final solution contains GO membranes sized between 3 and 8  $\mu$ m with a concentration of  $\sim 0.01$  wt%.

In static laser-scattering experiments, conformations of GO membranes are probed through the structure factor,  $S(q)$ , at a wave vector  $q$ , which is directly related to the radius of gyration,  $R_G$ , and to the fractal dimension<sup>9</sup>,  $d_f$ , of the scattering objects. The membrane conformations can be differentiated by their fractal dimensions: for a flat membrane,  $d_f = 2$ , whereas a compact conformation corresponds to  $d_f = 3$ . The fractal dimension of the intermediate crumpled phase is estimated to be 2.5, from a Flory argument<sup>3</sup>. For a monodisperse solution<sup>3,18</sup>,  $S(q) \approx 1 - (qR_G)^2/2$  for  $qR_G < 1$ , and  $S(q) \approx q^{-d_f}$  for  $qR_G > 1$ . As our solutions are polydisperse, the power law should be checked only at length scales shorter than the size of the smallest GO membranes (3  $\mu$ m). As our shortest probing length is 0.2  $\mu$ m (half of the laser wavelength), we have a range of about 1.5 decades in  $q$  to determine the scaling exponent  $d_f$  for the decay of  $S(q)$ .

Figure 2 illustrates the structure factors of GO membranes obtained from aqueous solutions at three pH values. The behaviour of  $S(q)$  is fairly consistent with a self-similar scaling of density fluctuations at length scales less than the size of membranes. The best fits give a value of  $d_f = 2.54 \pm 0.05$ , very close to the theoretical estimate<sup>3</sup> of 2.5. The fractal dimension was found to be insensitive to pH and salt concentrations. This is understandable considering the low density of the ionizable groups (one COOH group for every fifty carbons<sup>14</sup>). When 10% acetone is added to the solution, the structure factor, shown in Fig. 3, suggests formation of a compact conformation with a fractal dimension of  $\sim 3$ , and a somewhat smaller radius of gyration. This observation can be explained by the fact that acetone molecules are less polar than water molecules. On addition of acetone, the solvent becomes poorer, and the intra-membrane affinity is enhanced.

Although our observation of a fractal dimension of  $\sim 2.5$  is consistent with a crumpled phase of membranes, we also consider some other possible explanations. Haphazard crushing of

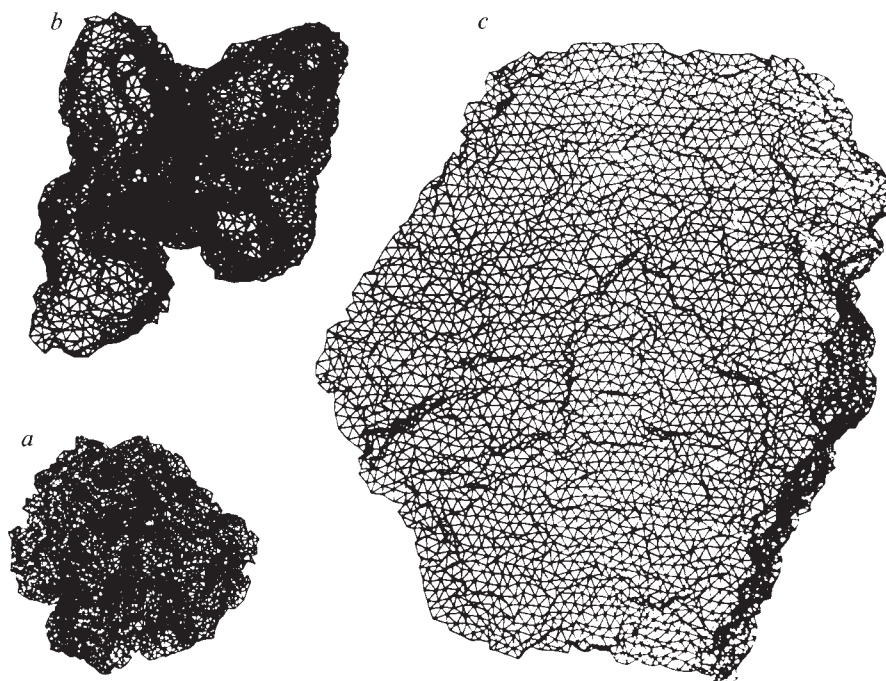


FIG. 1 Schematic representations of (a) compact, (b) crumpled and (c) flat conformations of solid membranes. These illustrations, taken from ref. 11, actually correspond to the same membrane at three different times in a numerical relaxation. This sequence also illustrates the need to complement static light-scattering results with dynamical studies.

foil or paper<sup>19,20</sup> also leads to objects with a fractal dimension of  $\sim 2.5$ . Such objects are, however, far from equilibrium and represent metastable states along the way to a fully compact conformation. The observation of a separate collapsed phase is evidence against this. Another possibility is a mixture of flat and compact configurations. This is unlikely in view of the lack of any curvature in the data of Fig. 2. On a log-log scale, light scattering data from a mixture of fractals of dimensions 2 and 3 would show clear bending in an interval of two decades

(D. A. Weitz, personal communication). Finally, flat membranes whose thickness scales with lateral size are inconsistent with our observations, as they would also produce a large curvature in  $\log q$ - $\log S(q)$  plots<sup>15</sup>.

Our experimental results therefore indicate that a crumpled conformation exists in solid membranes. This is somewhat surprising, as most numerical simulations of rather flexible self-avoiding membranes produce flat conformations<sup>4-6</sup>. The flat phase is, however, unstable to a glassy crumpled configuration in the presence of internal defects<sup>21</sup>. More information on the microstructure of GO, its rigidity and characteristic defects is needed. To determine the nature of the crumpled state, we propose three experimental studies. First, X-ray scattering experiments should be performed to reveal the mean curvatures and defects in GO membranes at microscopic scales. Second, the crumpled-to-compact transition should be systematically

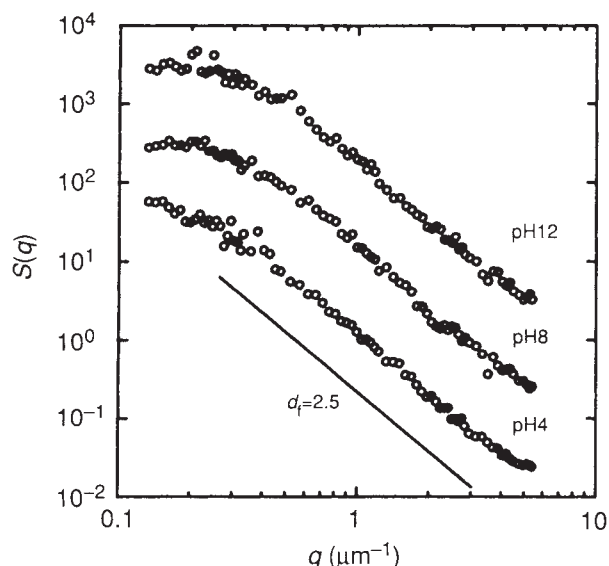


FIG. 2 The structure factor,  $S(q)$ , of GO membranes (sizes 3 to 8  $\mu\text{m}$ ) in aqueous solution at different pH. All measurements were made at room temperature. The wavevector is defined as  $q=1/\lambda$ , where  $\lambda$  is the wavelength of the density wave that scatters the incident laser beam. The data for pH 8 and pH 12 have been shifted vertically for clarity. For  $q > 0.33 \mu\text{m}^{-1}$ , the plots of  $\log S(q)$  against  $\log q$  can be fitted to straight lines, indicating that the GO membranes are self-similar at length scales smaller than their sizes. The fractal dimensions given by the fittings are:  $2.55 \pm 0.02$  for pH 4,  $2.56 \pm 0.03$  for pH 8, and  $2.51 \pm 0.03$  for pH 12. The theoretical estimate of a fractal dimension of 2.5 for a crumpled conformation is indicated for comparison by the solid line.

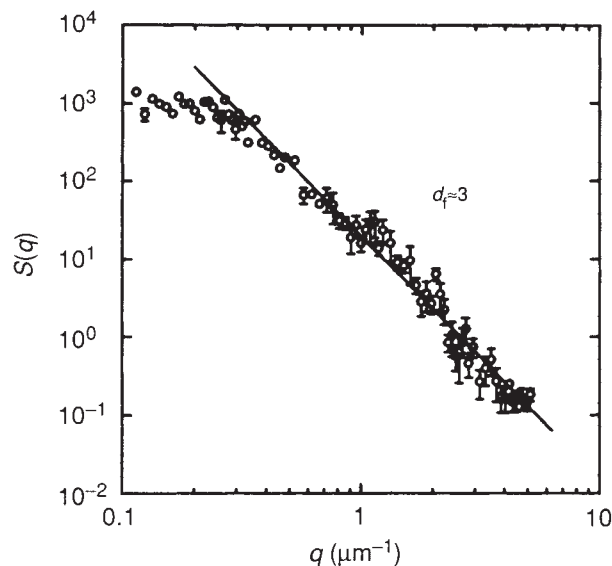


FIG. 3 The structure factor,  $S(q)$ , of GO membranes in a 10% acetone + water solution with pH 4. The slope at  $q > 0.33 \mu\text{m}^{-1}$  is  $\sim 3$  ( $3.10 \pm 0.05$ ), and the radius of gyration is somewhat smaller than in aqueous solutions, indicating a compact conformation of GO membranes.



studied with varying acetone concentrations and/or varying temperature at a suitable acetone concentration. A careful study of the reversibility of this transition could definitely test the existence of these two phases as distinct stable states. Third, a careful dynamic light-scattering study could reveal the relaxation times of the internal modes of the membranes, and thus distinguish between equilibrium and possible glassy conformations<sup>19–21</sup>. Preliminary dynamic light-scattering data show two dynamic modes, the slower of which has a relaxation time roughly proportional to  $q^{-2}$  and may correspond to brownian diffusion of the GO membranes. Further studies are in progress. □

Received 28 May; accepted 24 October 1991.

1. Lipowsky, R. *Nature* **349**, 475–481 (1991).
2. *Statistical Mechanics of Membranes and Surfaces* (eds Nelson, D. R., Piran, T. & Weinberg, S.) (World Scientific, Singapore, 1989).
3. Kantor, Y., Kardar, M. & Nelson, D. R. *Phys. Rev. Lett.* **57**, 791–794 (1986); *Phys. Rev. A* **35**, 3056–3071 (1987).
4. Plischke, M. & Boal, D. H. *Phys. Rev. A* **38**, 4943–4945 (1988).
5. Abraham, F. F., Rudge, W. E. & Plischke, M. *Phys. Rev. Lett.* **62**, 1757–1759 (1989).
6. Ho, J.-S. & Baumgartner, A. *Phys. Rev. Lett.* **63**, 1324 (1989).
7. Kantor, Y. & Nelson, D. R. *Phys. Rev. Lett.* **58**, 2774–2777 (1987).
8. Paczuski, M., Kardar, M. & Nelson, D. R. *Phys. Rev. Lett.* **60**, 2638–2640 (1988).
9. Mandelbrot, B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
10. Abraham, F. F. & Kardar, M. *Science* **252**, 419–422 (1991).
11. Abraham, F. F. & Nelson, D. R. *J. Phys. (France)* **51**, 2653–2672 (1990).
12. Brodie, B. *Phil. Trans.* **149**, 249–251 (1859).
13. Hummers, W. S. & Offeman, R. E. *J. Am. chem. Soc.* **80**, 1339 (1958).
14. Boehm, H. P., Glauss, A., Fischer, G. & Hofmann, U. *Z. Naturforsch.* **17**, 150–153 (1962).
15. Hwa, T., Kokufuta, E. & Tanaka, T. *Phys. Rev. A* **44**, R2235–R2237 (1991).
16. Chianelli, R. R., Prestidge, E. B., Pecoraro, T. A. & DeNeufville, J. P. *Science* **203**, 1105–1107 (1979).
17. Mutz, M., Bensimon, D. & Brienne, M. J. *Phys. Rev. Lett.* **67**, 923–926 (1991).
18. Martin, J. E. & Hurd, A. J. *J. appl. Cryst.* **20**, 61–78 (1987).
19. Gomes, M. A. F. & Vasconcelos, G. L. *Phys. Rev. Lett.* **60**, 237 (1988).
20. Kantor, Y., Kardar, M. & Nelson, D. R. *Phys. Rev. Lett.* **60**, 238 (1988).
21. Nelson, D. R. & Radzihovsky, L. *Europhys. Lett.* **16**, 79–84 (1991).

ACKNOWLEDGEMENTS. We thank D. R. Nelson, D. A. Weitz, T. A. Witten and Y. Q. Zhang for discussions. This research is supported by NSF grants to the CMSE at MIT.

## Magic numbers and stable structures for fullerenes, fullerides and fullerenium ions

Patrick W. Fowler\* & David E. Manolopoulos†

\* Department of Chemistry, University of Exeter, Exeter EX4 4QD, UK

† Department of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK

**MACROSCOPIC** amounts of the two fullerenes  $C_{60}$  and  $C_{70}$  have been available for a year<sup>1</sup>, and have already had an enormous impact on research in chemistry and physics. Experimentalists are now turning their attention to the higher fullerenes<sup>2,3</sup>. Qualitative molecular-orbital theory predicts<sup>4–6</sup> stability for  $C_n$  with  $n = 60, 70, (72), 76, 78, 84, \dots$ , of which all but  $C_{72}$  have now been produced by evaporation of graphite<sup>1–3</sup>, and in general for infinite series of closed-shell neutral fullerenes for  $n = 60 + 6k$  ( $k \neq 1$ ),  $70 + 30k$ ,  $84 + 36k$  (all  $k$ )<sup>7–9</sup>. Recent experimental observations of endohedral  $LaC_n$  metallofullerenes<sup>10</sup> have been rationalized in terms of 'magic numbers' for fulleride anions  $C_n^{2-}$ , for which special stability is predicted<sup>11</sup> at  $n = 74, 82, 88, \dots$ ; but the exact extent of charge transfer in these complexes has yet to be determined. Here we present calculations of magic numbers in the fullerenium sequence  $C_n^{2+}$  ( $n = 74, 80, (88), 92$ ) and show that the electron count determines stability and the atom count determines structure in all three (neutral, anionic and cationic) series. Stable cations have two carbons more, and stable anions two carbons less, than the corresponding stable neutral cluster. We predict likely structures of the 'magic' cations.

Qualitative theoretical treatments of carbon cages are based on the fullerene hypothesis: that stable structures are pseudo-spherical polyhedra with  $(n/2 + 2)$  faces, of which 12 are pen-

tagonal and the remainder hexagonal. The pentagons are needed for geometric closure, and the hexagons maximize the  $\pi$ -delocalization energy of these unsaturated molecules. Within the fullerene constraint, many isomeric forms of  $C_n$  are in principle possible, but stable structures have the following features in common: (1) isolated (non-adjointing) pentagons; (2) large  $\pi$ -bonding delocalization energies and large gaps between occupied and empty orbitals; and (3) low steric strain. The pentagon-isolation criterion has been elevated to the status of a rule (the IPR)<sup>12</sup>; it can be rationalized on steric grounds as a way of easing  $\sigma$ -strain<sup>13</sup>, and on electronic grounds as a way of avoiding the anti-aromatic eight-atom cycles inherent in pairs of adjacent pentagons<sup>4</sup>. Large delocalization energies and large gaps between the highest occupied and lowest unoccupied molecular orbitals are obviously conducive to electronic stability, and geometric (geodesic) stability of the cage framework is covered by criterion (3).

When  $n$  belongs to one of the special series  $60 + 6k$ ,  $70 + 30k$  or  $84 + 36k$ , candidate structures can be generated by the geometric leapfrog and carbon-cylinder constructions of refs 7–9. These constructions, of which one works by capping all faces and then taking the dual of a smaller fullerene, and the other by equatorial expansion of five- and six-fold symmetric fullerenes, automatically yield isolated-pentagon structures, and so satisfy the IPR. More generally, a computer search for all possible fullerene isomers of any given  $C_n$  cage can be carried out with the aid of the ring-spiral algorithm of ref. 14. This algorithm 'peels' each cage as a continuous spiral strip of edge-sharing pentagons and hexagons. The placing of the 12 pentagons in the sequence uniquely determines the adjacency matrix and hence the three-dimensional isomeric form of the fullerene. The three-dimensional structures are thus encoded in a compact one-dimensional representation. Structures with fused pentagons are easily detected and eliminated as the search proceeds. The general isomer count rises rapidly with  $n$  (from only 1 for  $n = 20$  to over 20,000 for  $n = 78$ , for example<sup>5</sup>). The count of structures with isolated pentagons also increases rapidly (from only 1 for  $n = 60$  to over 450 for  $n = 100$ ), but is much smaller for the same value of  $n$ .

Relative stabilities of isolated-pentagon structures for neutral fullerenes depend on a balance between the electronic and steric criteria mentioned above. The leapfrog and carbon cylinder series give all the fullerenes with formally closed-shell electronic structure, but other isomers may still have considerable electronic stability, albeit with slightly bonding LUMOs in simple Hückel theory. The electronic-steric balance is clearly a delicate one. Thus  $C_{76}$  is found<sup>3,4</sup> to be  $D_2$  (electronically favoured over the only other ( $T_d$ ) isolated-pentagon isomer), whereas<sup>5,15</sup> the predominant isomer of  $C_{78}$  is  $C_{2v}$  (sterically favoured over the four other isolated-pentagon isomers, one of which has a closed shell). Prediction of stability from molecular-orbital models or molecular mechanics<sup>15</sup> may therefore be a delicate matter. Nevertheless, steric strain effects being equal, leapfrog and

TABLE 1 Maximum energy gaps  $\Delta^{2+}$  for isolated-pentagon isomers of fullerenium cations  $C_n^{2+}$

| $n$ | $\Delta^{2+}$ | $n$ | $\Delta^{2+}$ |
|-----|---------------|-----|---------------|
| 70  | 0.0887        | 84  | 0.2259        |
| 72  | 0             | 86  | 0.2417        |
| 74  | 0.3371*       | 88  | 0.3121        |
| 76  | 0.1989        | 90  | 0.2138        |
| 78  | 0.2426        | 92  | 0.3147*       |
| 80  | 0.4195*       | 94  | 0.2906        |
| 82  | 0.2368        |     |               |

\* Local maximum in  $\Delta^{2+}$  corresponding to a single isomer that also has the highest  $\pi$ -bonding delocalization energy for the given value of  $n$ . Energies  $\Delta^{2+}$  are in units of a single effective Hückel resonance integral  $\beta$ .

TABLE 2 Electron counts for stable anionic, neutral and cationic fullerenes

| Electrons | 60       | 70       | 72            | 76            | 78            | 84            | 90            |
|-----------|----------|----------|---------------|---------------|---------------|---------------|---------------|
| anion     |          |          |               | $C_{74}^{2-}$ |               | $C_{82}^{2-}$ | $C_{88}^{2-}$ |
| neutral   | $C_{60}$ | $C_{70}$ | $(C_{72})$    | $C_{76}$      | $C_{78}$      | $C_{84}$      | $C_{90}$      |
| cation    |          |          | $C_{74}^{2+}$ |               | $C_{80}^{2+}$ |               | $C_{92}^{2+}$ |

Only  $\pi$ -electrons are included.

cylinder isomers are more likely to be stable than other neutral fullerenes, because they have the only properly closed electronic shells.

We can extend the discussion to fullerene anions  $C_n^{2-}$  and cations  $C_n^{2+}$  by modifying these arguments. As steric strain is largely independent of the charge on the cage, we must concentrate on the electronic criterion if we are to obtain any general 'first-order' result. It is also natural to look at addition and subtraction of pairs of electrons, because the singly charged species will always have open shells whereas doubly charged ions may have large HOMO-LUMO gaps. The critical energy gap is then  $\Delta^{2-}$  between the lowest and next-lowest unoccupied molecular orbitals of the neutral for  $C_n^{2-}$ , and  $\Delta^{2+}$  between the highest and next-highest occupied molecular orbitals of the neutral for  $C_n^{2+}$ . It seems reasonable that stable structures for these charged clusters should also be fullerenes with isolated pentagons, because other connectivities or adjacent pentagons would weaken the  $\sigma$ -framework, and other ring sizes would

detract from the  $\pi$ -stabilization. But as the requirements for the energy gap are different, different magic numbers and favoured isomers are to be expected for anionic, cationic and neutral species.

A search of isolated-pentagon fullerene isomers for di-anions in the range from 60 to 90 atoms has been reported<sup>11</sup>. Local maxima in  $\Delta^{2-}$  and delocalization energy suggest that  $C_{74}^{2-}$ ,  $C_{82}^{2-}$  and  $C_{88}^{2-}$ , will be more stable than their neighbours. Assuming equal steric strain, these cages are predicted to adopt structures with  $D_{3h}$ ,  $C_{3v}$  and  $C_2$  symmetry<sup>11</sup>, respectively. The basic trend in the HOMO-LUMO gap, however, is largely independent of any definite structural proposal for a given  $C_n^{2-}$ .

The most stable  $C_n^{2-}$  cages have been proposed as partners in endohedral metal complexes of the type  $MC_n$  (where M is Ba, La and so on) in which the metal atom M is expected to donate two valence electrons to the  $\pi$ -system of the enclosing carbon cage. The predicted stability of  $C_{3v} C_{82}^{2-}$  may account for the experimental observation of a long-lived  $LaC_{82}$  species<sup>10</sup>, given that the two available electronic structure calculations on  $LaC_{60}$ <sup>16,17</sup> are consistent with formal charge distributions  $La^{\delta+}C_n^{\delta-}$ , where  $\delta$  lies somewhere between 2 and 3. Electron spin resonance studies<sup>18</sup> suggest that the extent of electron transfer to the cage may be greater than assumed in the qualitative Hückel treatment (that is,  $\delta > 2$ ), but the observation of the characteristic La nuclear-spin splitting pattern also indicates residual spin-density at La ( $\delta < 3$ ). X-ray photoelectron spectroscopy measurements on  $LaC_n$  species<sup>19</sup> also support a value of  $\delta$  greater than two, and for cages with a triply degenerate LUMO,  $\delta$  is likely to approach the upper limit of 3. The precise value is not important for qualitative prediction.

A similar search of isomeric fullerenium cations reveals a different pattern. Whereas closed-shell anions are common in the range 70-90, no cation examined has a properly closed shell (this would require a neutral with an unlikely antibonding HOMO). Nevertheless, as Table 1 shows, there are distinct maxima in  $\Delta^{2+}$  at  $n = 74, 80$  and  $92$ . Each of these isomers also has the highest  $\pi$ -bonding delocalization energy for the given value of  $n$ . A maximum in  $\Delta^{2+}$  occurs at  $n = 88$ , but not for the isomer of maximal delocalization energy, and so the situation for  $C_{88}^{2+}$  is less clear cut. Figure 1 shows the structures of the 'best' cations  $C_{74}^{2+}$ ,  $C_{80}^{2+}$  and  $C_{92}^{2+}$ , along with two possible structures of  $C_{88}^{2+}$ . The 74-atom cage coincides with the structural proposal for the  $C_{74}^{2-}$  anion. The 80-atom structure is at a clear local maximum in  $\Delta^{2+}$  and would seem to be the best candidate for a stable cation on electronic grounds of all those examined. There is also an icosahedral  $C_{80}$  cage which has a four-fold-degenerate HOMO containing only two electrons, as do all icosahedral cages with  $60m + 20$  atoms<sup>20</sup>, but the HOMO-LUMO gap of the corresponding  $C_{80}^{2+}$  cation is only  $0.3441\beta$ , which is beaten by the  $D_{5d}$  isomer in Fig. 1. Nevertheless, because  $I_h$  symmetry gives the least strain in the  $\sigma$ -framework (criterion (3)), it will be interesting to see whether more sophisticated modelling alters the predicted isomer preference for  $C_{80}^{2+}$ . The icosahedral isomer is illustrated in Fig. 2.

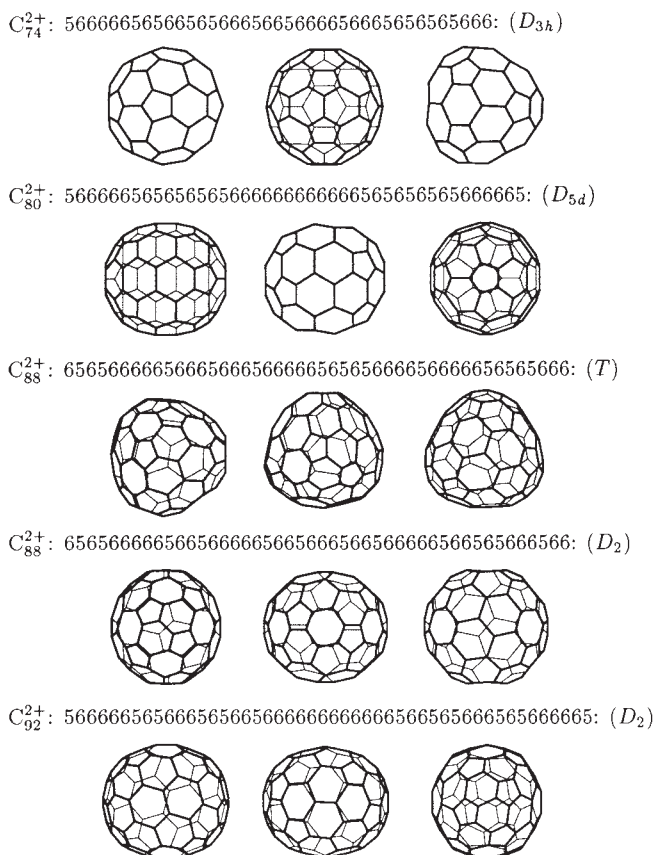


FIG. 1 Cage structures of fullerenium cations predicted by qualitative molecular orbital theory assuming equal steric strain. Views down three orthogonal axes are shown as an aid to model building. For the 74-, 80- and 92-atom cages there is a single 'best' structure, but for the 88-atom cage two isomers are shown. The first has the larger band gap ( $0.3121\beta$  against  $0.1850\beta$ ), the second a slightly better delocalization energy ( $0.5857\beta$  against  $0.5846\beta$ ). A ring-spiral formula<sup>14</sup> is given for each isomer.

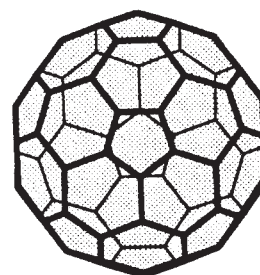


FIG. 2 The icosahedral 80-atom fullerene cage. This isomer of  $C_{80}^{2+}$  would have the best distribution of  $\sigma$  strain, but is less favoured on  $\pi$  electronic grounds than the  $D_{5d}$  isomer in Fig. 1.



Setting aside the ambiguous case of  $C_{88}^{2+}$ , the outline of a general pattern for anions and cations emerges. The stable electron counts are given in Table 2, from which the significance of the shell closure idea<sup>10</sup> is apparent. Cations drop down to the electron count of a stable neutral, and anions step up. But the geometrical structures are determined by the atom count. Within the isolated-pentagon restriction there is only one 74-atom cage and so the anion and cation are isostructural, but in general different atom counts and structures are found for each series.

Shell structures are common in physics and chemistry<sup>21</sup>. Many-fermion systems with roughly spherical symmetry, for which the independent-particle approximation gives an effective central potential, show shells governed by effective spherical quantum numbers. For the fullerenes, we have observed shell closure by considering all possible bonding topologies within the qualitative molecular-orbital model; a 'spherical' model of the TSH type<sup>22</sup> which does not invoke any particular topology and yields a shell structure directly is presumably also possible.

The boranes  $B_nH_m^q$  and transition metal clusters show a similarly strong link between atom/electron counts and structure<sup>23</sup>. In the electron-deficient boranes, the bonding is many centred, and addition of electrons causes some triangular faces of the *closo*  $B_nH_n^{2-}$  polyhedron to open out into squares or pentagons to give the *nido*  $B_nH_n^{4-}$  series. Our model for charged fullerenes precludes ring expansion for the reasons set out earlier, but does allow the shape of a given *n*-atomic cage to depend on the net charge on the cluster.

Cationic fullerenes have been invoked in several laser-vaporization and electrochemical experiments, but may occur more widely. Encapsulation of an electronegative element could lead to 'internal-zwitterion' counterparts of those endohedral metallofullerene complexes in which an anionic cage contains a cationic metal<sup>10,11</sup>. Atoms such as fluorine and oxygen seem to attack the cage rather than enter it<sup>2,24</sup>, but many more elements have yet to be tried. Doping of boron or nitrogen<sup>25</sup> into the framework of a charged carbon cluster could also produce one of the stable electron counts. It is suggestive that  $C_{74}^{2+}$  has a symmetry-equivalent pair of antipodal sites; occupation of these by two boron atoms could give the best structure of  $C_{72}B_2$  retaining the  $D_{3h}$  symmetry of the 'magic'  $C_{74}^{2+}$  cage. Mixing electrophilic and nucleophilic cages in a lattice structure, one can envisage hypothetical ionic forms of solid carbon, fullerenes and fullerene salts, or even 'solvated' forms of a salt MX in which the cations lie inside one set of cages and the anions in the other. Our present results suggest that a search for  $C_{80}^{2+}$ , or an analogue such as  $C_{78}B_2$ , would be a fruitful direction for future experiment. □

## Multiple phases of polymer gels

Masahiko Annaka\* & Toyochi Tanaka

Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

\* Permanent address: Mitsubishi Kasei Corporation, Yokohama, Japan

SYNTHETIC polymer gels are known to exist in two phases, swollen and collapsed<sup>1,2</sup>. Volume transitions may occur between the phases either continuously or discontinuously. We report here that more than two phases can be found in gels consisting of copolymers of randomly distributed positively and negatively charged groups. In these gels, polymer segments interact with each other through attractive or repulsive electrostatic interactions and through hydrogen bonding. It is the combination of these forces that seems to result in the existence of several phases. Each phase is characterized by a distinct degree of swelling, with abrupt jumps between them. The number of phases depends on the proportions of positively and negatively charged monomers, and decreases from a maximum of seven to just one for pure cationic or anionic gel compositions. The existence of these phases presumably reflects the ability of macromolecular systems to adopt different stable conformations in response to changes in environmental conditions.

We prepared copolymer gels of acrylic acid (the anionic constituent) and methacryl-amido-propyl-trimethyl ammonium chloride (cationic, MAPTAC) by free radical polymerization<sup>3</sup>. Their molar ratio was varied while the total molar amount was fixed at 700 mmol. For example, 350 mmol of acrylic acid, 350 mmol of MAPTAC, 0.133 g of N,N'-methylene bisacrylamide (as a cross-linker) and 40 mg of ammonium persulphate (the initiator) were dissolved in 100 ml of water. The solution was degassed and polymerized in capillaries of inner diameter  $d_0 = 0.34$  mm at 60 °C. An acrylic acid molecule can form hydrogen bonds with other acrylic acid molecules, and can attract MAPTAC electrostatically (Fig. 1).

The gel was cut into a small cylinder of length 1 mm and placed in a glass cell. The gel was continuously flushed with water from a reservoir, in which the pH was controlled by adding hydrochloric acid solution (to lower the pH) or water (to increase the pH) below pH 7. To obtain a pH higher than pH 7, sodium hydroxide solution was used. The equilibrium gel diameter,  $d$ , was measured under a microscope.

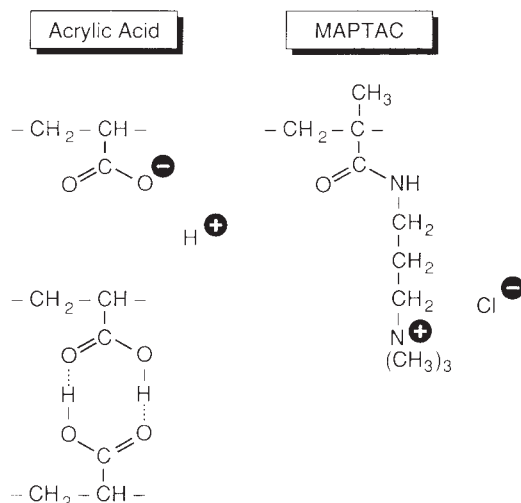


FIG. 1 Chemical structures of acrylic acid and MAPTAC. Acrylic acid is a weak acid ionized at  $pH > 4.5$ , and can form hydrogen bonds to another acrylic acid. MAPTAC is a strong base and considered to be ionized in most of the pH range of our measurements.

Received 21 October; accepted 26 November 1991.

- Krätschmer, K., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
- Diederich, F. *et al. Science* **252**, 548–551 (1991).
- Ettl, R., Chao, I., Diederich, F. & Whetten, R. L. *Nature* **353**, 149–153 (1991).
- Manolopoulos, D. E. *J. chem. Soc. Faraday Trans.* **87**, 2861–2862 (1991).
- Fowler, P. W., Manolopoulos, D. E. & Batten, R. C. *J. chem. Soc. Faraday Trans.* **87**, 3103–3104 (1991).
- Fowler, P. W. *J. chem. Soc. Faraday Trans.* **87**, 1945–1946 (1991).
- Fowler, P. W. & Steer, J. I. *J. chem. Soc. Chem. Commun.* 1403–1405 (1987).
- Fowler, P. W., Cremona, J. E. & Steer, J. I. *Theor. Chim. Acta* **73**, 1–26 (1988).
- Fowler, P. W. *J. chem. Soc. Faraday Trans.* **86**, 2073–2077 (1990).
- Chai, Y. *et al. J. phys. Chem.* (in the press).
- Manolopoulos, D. E. & Fowler, P. W. *Chem. Phys. Lett.* **187**, 1–7 (1991).
- Haefliger, R. E. *et al. Mat. Res. Soc. Proc.* **206**, 627–638 (1991).
- Kroto, H. W. *Nature* **329**, 529–531 (1987).
- Manolopoulos, D. E., May, J. C. & Down, S. E. *Chem. Phys. Lett.* **181**, 105–111 (1991).
- Diederich, F. *et al. Science* (in the press).
- Rosen, A. & Wastberg, B. *J. Am. chem. Soc.* **110**, 8701–8703 (1988).
- Chang, A. H. H., Ermler, W. C. & Pitzer, R. M. *J. chem. Phys.* **94**, 5004–5010 (1991).
- Johnson, R. D., de Vries, M. S., Salem, J., Bethune, D. S. & Yannoni, C. S. *Nature* (in the press).
- Weaver, J. H. *et al. Chem. Phys. Lett.* (in the press).
- Fowler, P. W. *Chem. Phys. Lett.* **131**, 444–450 (1986).
- de Heer, W. A., Knight, W. D., Chou, M. Y. & Cohen, M. L. *Solid State Phys.* **40**, 93–181 (1981).
- Stone, A. J. *Inorg. Chem.* **20**, 563–571 (1981).
- Wade, K. J. *J. chem. Soc. Chem. Commun.* 792–793 (1971).
- Holloway, J. H. *et al. J. chem. Soc. Chem. Commun.* 966–969 (1991).
- Guo, T., Jin, C. & Smalley, R. E. *J. phys. Chem.* **95**, 4948–4950 (1991).

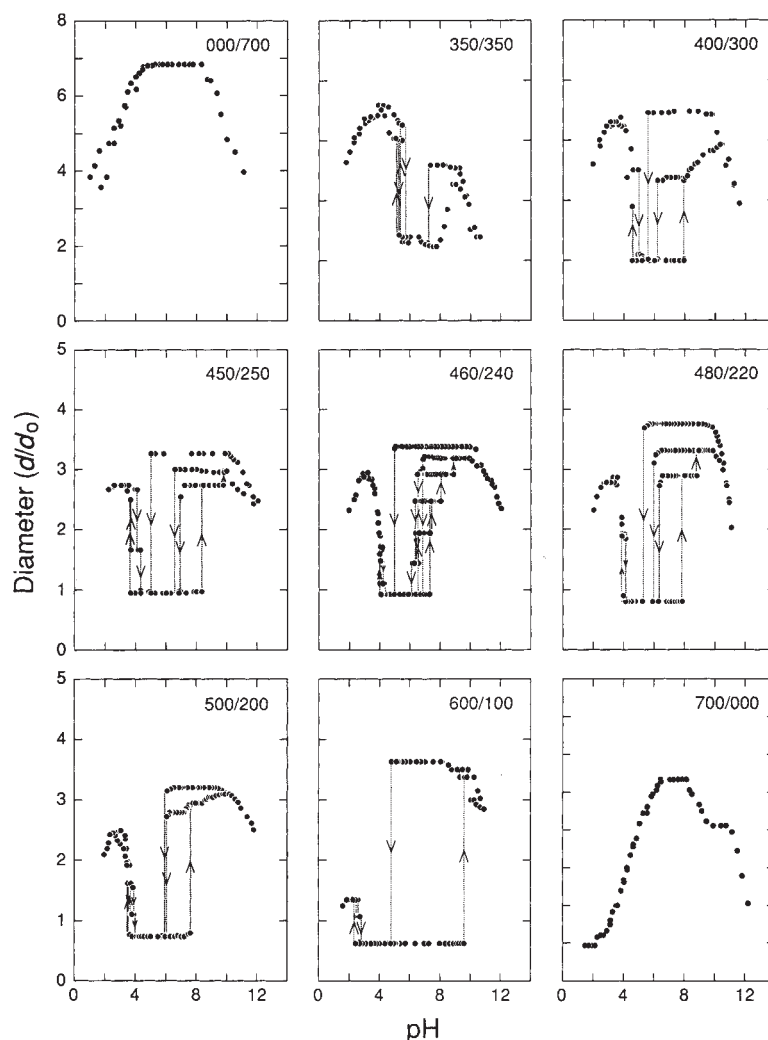


FIG. 2 Equilibrium swelling degree,  $d/d_0$ , of copolymer gels, made of various molar ratios of acrylic acid and MAPTAC in water, as a function of pH at 25 °C. Acrylic acid and MAPTAC ratios of the gels are indicated. For example, 450/250 means that the gel consists of 450 mmol of acrylic acid and 250 mmol of MAPTAC. For some gels there appear multiple phases for the same value of pH. These are reached by varying pH through different paths. See text for details of the pathways followed for the 450/250 gels. The swelling curves for other gels are obtained in a similar way. All the hysteresis cycles were repeated with excellent reproducibility.

Figure 2 shows the swelling curves of copolymer gels with different compositions. (Similar gels have been studied previously, but no phase transition was observed<sup>4</sup>.) Consider the gel labelled 450/250, which consists of 450 mmol acrylic acid and 250 mmol MAPTAC. At neutral pH, the gel had a diameter  $d = d_0$ , and we refer to this as phase 1.0 ( $d/d_0 = 1$ ). As the pH was increased the gel swelled discontinuously to phase 2.7 at pH 8.5. If the pH was lowered from 8.5, the gel collapsed back to phase 1.0 at pH 7.0. If, instead, the pH was further increased from 8.5, the gel swelled discontinuously into phase 3.0 at pH 9.8. As pH was lowered from 9.8 the gel collapsed into phase 1.0 at pH 6.4. If, instead, the pH was increased to 12, the gel diameter changed continuously, and when the pH was lowered again, it swelled continuously to phase 3.3. Further reduction of the pH caused the gel to collapse to phase 1.0 at pH 4.9. These cycles were repeated twice or more with excellent reproducibility. Three phases, 1.0, 1.7 and 2.6, were found at lower pH.

A systematic change in the macroscopic phase behaviour was observed when the polymer composition was varied gradually. The largest number of phases was seven for the 460/240 gel, and decreased as the ratio deviated from the value. When temperature was varied for a fixed pH, the gel underwent discontinuous transitions between some of the phases confirmed in the pH experiments (Fig. 3). Note that none of the shrunken phases shown in Fig. 2 is the most compact phase, which was found by adding acetone, a nonpolar solvent (Fig. 4).

We monitored the kinetics of swelling and shrinking under different conditions. Characteristic relaxation times of the gels in response to changes in temperature, pH or solvent are minutes for continuous changes and several hours for large discontinuous changes.

The onset of a discontinuous transition, however, was readily noticed within several seconds, as a quick change in gel diameter. Selected phases were confirmed to be stable for more than a month.

Multiple phases can coexist if each corresponds to a local minimum in free energy. The lowest minimum is the stable equilibrium state, however, and the others represent metastable

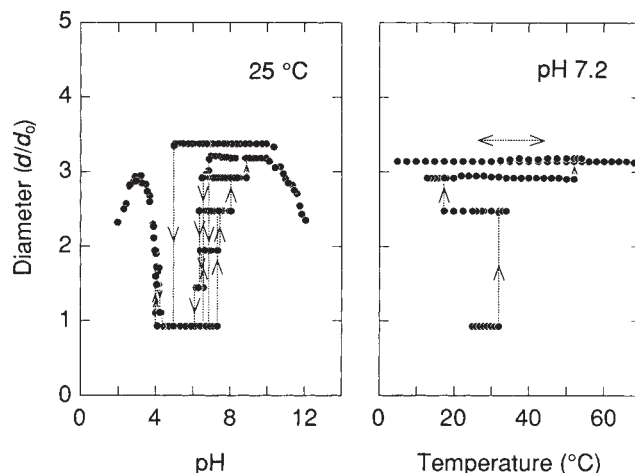
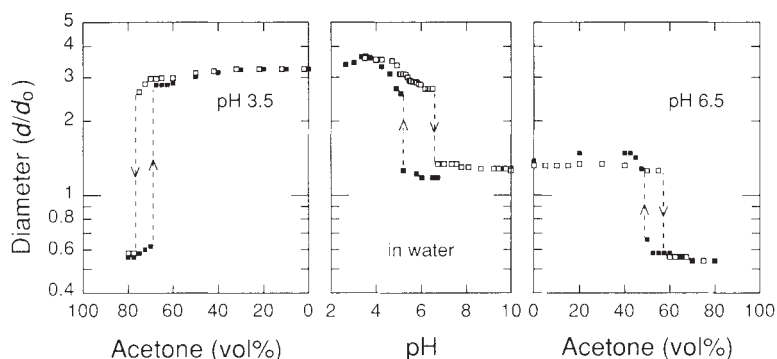


FIG. 3 Equilibrium swelling degree  $d/d_0$  of copolymer gel made of 460 mmol of acrylic acid and 240 mmol of MAPTAC as a function of temperature at pH 7.2 (right), and pH dependence (left) taken from Fig. 2.



FIG. 4 Centre: the equilibrium diameter of the copolymer gel of 350 mmol acrylic acid and 350 mmol MAPTAC in water as a function of pH. The gel is in phase 1.2 at neutral pH. As the pH is lowered (by adding HCl), it swells discontinuously at pH 5.1, and shrinks discontinuously at pH 6.5 when the pH is raised again by adding water. Right: the gel diameter in acetone solutions of neutral pH as a function of acetone concentration (vol% before mixing). A discontinuous volume transition and associated hysteresis is observed between the intermediate phase and the collapsed phase. Left: the gel diameter in acetone solutions at pH 3.5 as a function of acetone concentration. A discontinuous volume transition is observed between the collapsed and the swollen phases.



phases. As the environment varies, the minima can cross, leading to a discontinuous phase transition. The fact that observed transitions always involve a discrete set of swelling degrees and none of the intermediate values, suggests that these swelling degrees correspond to separate free-energy minima.

Theoretically, multiple phases may be understood as a result of competition among various factors. The mean-field free energy consists of terms for rubber elasticity, osmotic pressure by counter ions, net charge repulsion and virial interactions<sup>4-6</sup>. The effect of having equal amounts of cationic and anionic charges on polymers, the polyampholyte effect, results in long-range attractive interactions due to the Debye-Hückel energy<sup>7-10</sup>, and in short-range repulsive interactions due to local charge imbalance<sup>11</sup>. These six or seven terms with different powers of polymer density can create free-energy minima at three distinct densities. It is therefore necessary to introduce new order parameters (variables that characterize the transitions) in addition to the polymer density to predict more than three phases.

The hydrogen bonding density,  $\rho_{HB}$ , may be a natural choice for a new order parameter. Formation of hydrogen bonding is energetically favourable, but entropically undesirable as it restricts the freedom of chain configurations. This competition can create two free-energy minima for a fixed polymer density  $\rho$ . As  $\rho_{HB}$  and  $\rho$  are coupled, the free energy,  $F(\rho, \rho_{HB})$ , can have six minima in the  $\rho$ - $\rho_{HB}$  plane. We are now developing a theoretical description of the transitions in which  $\rho_{HB}$  is the order parameter.

It is important to determine the local polymer order, if any, at each phase. As the polymers in a gel are randomly cross-linked and finite in length, there should be no symmetry breaking of the gel as a whole. Because of the competing interactions, some of the phases, in particular the denser phase, may be non-ergodic<sup>12-14</sup>. That is, the gel is trapped in a configuration corresponding to a local free-energy minimum and cannot experience all the other possible configurations with the same free energy. In this case different cycles may trap the gel in different free-energy minima.

Carbon-13 NMR was done on the 480/220 gel in which three carbons of the acrylic acid are replaced by the isotope <sup>13</sup>C. The NMR spectra for different phases at the same pH were clearly distinguishable, indicating that each phase has a different local environment. More extensive study is needed, however, to identify, from these NMR or other data, the microscopic structure of the new phases. □

11. Kantor, Y. & Kardar, M. *Europhys. Lett.* **14**, 421-426 (1991).
12. Golubovic, L. & Lubensky, T. *Phys. Rev. Lett.* **63**, 1082-1085 (1989).
13. Shakhnovich, E. I. & Gutin, A. M. *Europhys. Lett.* **8**, 327-332 (1989).
14. Goldbart, P. & Goldfeld, N. *Phys. Rev. Lett.* **58**, 2676-2679 (1987).

ACKNOWLEDGEMENTS. We thank M. Kardar, M. Shibayama, T. Tokuihoro, J. F. Joanny, A. R. Khokhlov, A. Yu. Grosberg and K. Wasserman for discussions, and D. Western for technical assistance. M.A.'s stay at MIT was supported by Mitsubishi Kasei Co. This work was supported by the NSF.

## Observation of hypervalent $\text{CLi}_6$ by Knudsen-effusion mass spectrometry

Hiroshi Kudo

Japan Atomic Energy Research Institute, Tokai-mura, Ibaraki-ken 319-11, Japan

**BECAUSE** of the unusual geometry, bonding and reactivity of some lithiated molecules, they (polylithiated hydrocarbons in particular) have attracted interest from organic, organometallic and theoretical chemists<sup>1-7</sup>. Calculations by Schleyer *et al.*<sup>3</sup> have predicted that the hypervalent carbon-lithium molecules  $\text{CLi}_5$  and  $\text{CLi}_6$  should be stable with respect to dissociation to  $\text{CLi}_4$  and  $\text{Li}_2$ . Here I report the observation of  $\text{CLi}_6$ , which forms in the gas phase over solid  $\text{Li}_2\text{C}_2$  and is identified by Knudsen-effusion mass spectrometry<sup>8</sup>. I confirm that the dissociation reaction above is significantly endothermic, indicating that carbon can expand its octet of electrons to form this relatively stable molecule.

The molecular structures of  $\text{CLi}_5$  and  $\text{CLi}_6$  proposed by Schleyer *et al.*<sup>3</sup> (Fig. 1) suggest that the carbon atoms in these molecules may be surrounded by nine and ten valence electrons, violating, at least formally, the octet rule; elements in the second row of the periodic table normally prefer to be surrounded by eight electrons. These exotic molecules are called hyperlithiated or hypervalent molecules.

There is increasing evidence for the existence of hyperlithiated compounds. Lagow and coworkers synthesized solid products such as  $(\text{CLi}_4)_n$  by the reaction of lithium vapour with  $\text{CCl}_4$ , and observed lithiated pentacoordinate carbocations (such as  $\text{CLi}_5^+$ ,  $\text{CHLi}_4^+$  and  $\text{CH}_3\text{Li}_3^+$ ) by flash-vaporization mass spectrometry<sup>9,10</sup>. Wu and Ihle observed  $\text{CLi}_n$  species in the lithium vapour diffusing out of graphite membranes at high temperatures<sup>11</sup>. Schmidbaur *et al.*<sup>12-14</sup> made structural analyses and spectroscopic studies on hypervalent polyauroalkane cations containing penta- and hexacoordinate carbon atoms. The dissociation energies of hyperlithiated molecules have been experimentally determined for only a few molecules such as  $\text{Li}_3\text{O}$ ,  $\text{Li}_4\text{O}$ ,  $\text{Li}_5\text{O}$ ,  $\text{Li}_3\text{S}$  and  $\text{Li}_4\text{S}$  (refs 15-18).

I have observed molecular species in the gas phase over  $\text{Li}_2\text{C}_2(\text{s})$  by means of Knudsen-effusion mass spectrometry, and have previously<sup>19</sup> reported the dissociation energies of  $\text{CLi}_3$  and  $\text{CLi}_4$ . Here I report the dissociation and atomization energies of  $\text{CLi}_6$  observed in the vapour over  $\text{Li}_2\text{C}_2(\text{s})$ .

Received 25 February; accepted 5 November 1991.

1. Dusek, K. & Patterson, D. J. *Polym. Sci. A2* **6**, 1209-1216 (1968).
2. Tanaka, T. *et al. Phys. Rev. Lett.* **45**, 1636-1639 (1980).
3. Ilmain, F., Tanaka, T. & Kokufuta, E. *Nature* **349**, 400-401 (1991).
4. Myoga, A. & Katayama, S. *Polym. Prep. Japan* **36**, 2852-2853 (1987).
5. DeGennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell University, Ithaca, New York, 1979).
6. Edwards, S., King, P. R. & Pincus, P. *Ferroelectrics* **30**, 3-6 (1980).
7. Lifshitz, I. M., Grosberg, A. Yu & Khokhlov, A. R. *Rev. mod. Phys.* **50**, 683-713 (1978).
8. Higgs, P. G. & Joanny, J.-F. *J. chem. Phys.* **94**, 1543-1554 (1991).
9. Qian, C. & Khokhlov, A. L. *J. chem. Phys.* **89**, 5273-5279 (1988).
10. Khokhlov, A. R. & Kachaturian, K. A. *Polymer* **23**, 1742-1750 (1982).

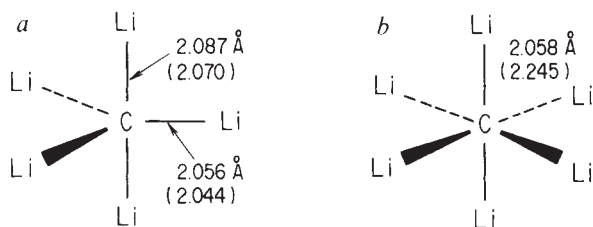


FIG. 1 Molecular structures of (a)  $\text{CLi}_5$  (symmetry group  $D_{3h}$ ) and (b)  $\text{CLi}_6$  (symmetry group  $O_h$ ) calculated by Schleyer *et al.* with the 3-21G basis set<sup>3</sup> and (values in parentheses) for 6-31G\* (P. v. R. Schleyer, personal communication).

Knudsen-effusion mass spectrometry is used widely in thermodynamic studies to measure equilibrium partial pressures of gaseous species over liquid and solid materials. Details of the experimental apparatus and procedures have been described previously<sup>19</sup>; briefly, they involve the measurement of mass spectra of molecular beams effusing from a Knudsen cell at high temperatures. Crystalline  $\text{Li}_2\text{C}_2$  (99% pure, Cerac/Pure Inc.) loaded in a molybdenum Knudsen cell was placed in an ultra-high-vacuum system equipped with a quadrupole mass spectrometer (Extrel C-80). The volume of the Knudsen cell was  $1.5\text{ cm}^3$  and the orifice diameter was 0.3 mm. The cell was heated by a radiofrequency generator (JEOL; 5 kW, 0.2 MHz), the temperature being controlled within  $\pm 5\text{ K}$ . Molecular species effusing from the cell were directly introduced into a cross-beam ionizer. A shutter was installed between the Knudsen cell and the ionizer to discriminate the molecular beam from the residual gas; the beam was analysed by the mass spectrometer only when the shutter was open.

Mass spectra of the molecular species effusing from the cell at 1,052 K are shown in Fig. 2. The mass peaks at mass/charge ratios  $m/e = 52, 53$  and  $54$  are the signals arising from different isotopic combinations in  $\text{CLi}_6$ ;  $m/e = 52$  for  $^{12}\text{C}^{67}\text{Li}_2^{7}\text{Li}_4$ <sup>+</sup> and  $^{13}\text{C}^{67}\text{Li}_3^{7}\text{Li}_3$ <sup>+</sup>,  $m/e = 53$  for  $^{12}\text{C}^{67}\text{Li}^{7}\text{Li}_5$ <sup>+</sup> and  $^{13}\text{C}^{67}\text{Li}_2^{7}\text{Li}_4$ <sup>+</sup>, and  $m/e = 54$  for  $^{12}\text{C}^{7}\text{Li}_6$ <sup>+</sup> and  $^{13}\text{C}^{67}\text{Li}^{7}\text{Li}_5$ <sup>+</sup>. The observed pattern coefficients of these signals (0.1, 0.5 and 1.0) agreed with that calculated for a  $\text{CLi}_6$  molecule consisting of one carbon atom and six lithium atoms with the natural isotopic abundance (7.5%  $^6\text{Li}$ , 92.5%  $^7\text{Li}$ , 98.9%  $^{12}\text{C}$ , 1.1%  $^{13}\text{C}$ ). In addition, these signals were detected only when the shutter was open, indicating that these species effused from the cell. Although the ionization potential of  $\text{CLi}_6$  was not determined, because of insufficient signal intensity, the signals due to  $\text{CLi}_6$  disappeared when the energy of the impact electrons was lowered below 9 eV or the emission current was turned off. These facts indicate that the  $\text{CLi}_6$  species is a neutral molecule existing in the gas phase over  $\text{Li}_2\text{C}_2(\text{s})$ . No distinct signals were observed for another interesting molecule,  $\text{CLi}_5$ . It is possible that the  $\text{CLi}_5$  molecule, although energetically stable, is not necessarily stable towards

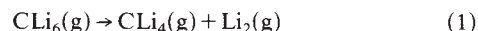
TABLE 1 Calculation of enthalpy for the reaction  $\text{CLi}_6(\text{g}) \rightleftharpoons \text{CLi}_4(\text{g}) + \text{Li}_2(\text{g})$

| $T$ (K)  | $-R \ln K_p$ | $-\Delta[(G_T^\circ - H_{298}^\circ)/T]$<br>( $\text{J mol}^{-1} \text{K}^{-1}$ ) | $\Delta H_{298}^\circ$<br>( $\text{kJ mol}^{-1}$ ) |
|----------|--------------|-----------------------------------------------------------------------------------|----------------------------------------------------|
| 961      | 130.6        | 146.9                                                                             | 266.7                                              |
| 973      | 136.4        | 146.8                                                                             | 275.6                                              |
| 987      | 138.2        | 146.7                                                                             | 281.2                                              |
| 1,052    | 112.7        | 146.3                                                                             | 272.5                                              |
| 1,075    | 112.4        | 146.1                                                                             | 277.9                                              |
| 1,083    | 103.5        | 146.1                                                                             | 270.3                                              |
| 1,168    | 95.6         | 145.6                                                                             | 281.7                                              |
| Average* |              |                                                                                   | $275.1 \pm 11.2$                                   |
|          |              |                                                                                   | $\Delta H_0^\circ = 271.6 \pm 11.2$                |

\* The uncertainty quoted is standard deviation ( $2\sigma$ ).

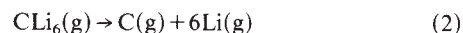
association or reactions with other molecules. The partial pressure  $p_i$  of the  $i$ th species in the Knudsen cell was obtained from the relation<sup>8</sup>  $p_i = \kappa [I_i T / (\sigma_i \beta_i \gamma_i)]$ , where  $\kappa$  is the proportionality constant,  $I_i$  the signal intensity,  $\sigma_i$  the relative ionization cross-section,  $\beta_i$  the isotope abundance and  $\gamma_i$  the gain of the detector in the mass spectrometer. The equilibrium partial pressures of gaseous species over  $\text{Li}_2\text{C}_2(\text{s})$  are shown in Fig. 3. The proportionality constant  $\kappa$  was determined on the basis of equilibrium of the  $\text{Li}_2(\text{g}) \rightleftharpoons 2\text{Li}(\text{g})$  process<sup>20</sup>. The molecular ionization cross-section was calculated by taking the sum of Mann's atomic cross-sections<sup>21</sup>. The gain of the detector (an electron multiplier) was obtained from a calibration curve.

By adopting the partial pressures of  $\text{CLi}_6(\text{g})$ ,  $\text{CLi}_4(\text{g})$  and  $\text{Li}_2(\text{g})$  observed in the range 961–1,168 K, the equilibrium constants  $K_p = p_{\text{CLi}_4} p_{\text{Li}_2} / p_{\text{CLi}_6}$  for reaction (1)



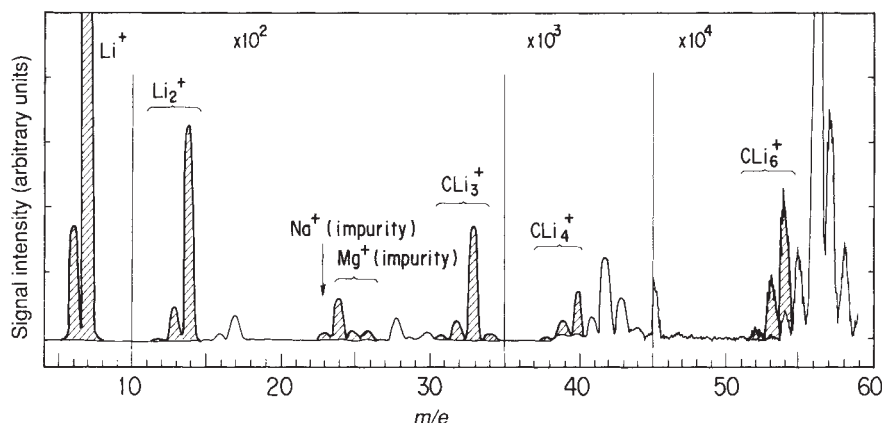
were calculated, and then the reaction enthalpy was evaluated from the relation  $-\Delta H_{298}^\circ / T = R \ln K_p + \Delta[(G_T^\circ - H_{298}^\circ) / T]$ , as listed in Table 1. Here, the free energy functions,  $(G_T^\circ - H_{298}^\circ) / T$ , for  $\text{Li}(\text{g})$  and  $\text{Li}_2(\text{g})$  were taken from ref. 22. The free energy function for  $\text{CLi}_6(\text{g})$  as well as for  $\text{CLi}_4(\text{g})$  was calculated by a statistical dynamical method, using molecular parameters given by P. v. R. Schleyer *et al.* (personal communication). The resulting enthalpy for reaction (1),  $\Delta H_0^\circ = 272 \pm 11\text{ kJ mol}^{-1}$ , agrees well with the theoretical value,  $\Delta H_0^\circ = 273\text{ kJ mol}^{-1}$ , reported by Schleyer *et al.*<sup>3</sup>; at higher theoretical levels of *ab initio* calculations with the 6-31G\* basis set, they calculate the enthalpy for reaction (1) to be  $275\text{ kJ mol}^{-1}$ , confirming the earlier result (P. v. R. Schleyer, personal communication).

Furthermore, the enthalpy for the reaction



was evaluated by combining  $\Delta H_{298}^\circ = 832 \pm 32\text{ kJ mol}^{-1}$ , obtained for the reaction  $\text{CLi}_6(\text{g}) \rightarrow \text{C}(\text{s}) + 6\text{Li}(\text{g})$  from the observed partial pressures of  $\text{CLi}_6(\text{g})$  and  $\text{Li}(\text{g})$ , and the published

FIG. 2 Mass spectrum of gaseous species in molecular beams effusing from the Knudsen cell containing  $\text{Li}_2\text{C}_2(\text{s})$  at 1,052 K. Different regions of the spectrum are shown at different intensity scales. The gaseous species were ionized by electron impact at 13.0 eV, an energy low enough to prevent the fragmentation of parent species. The bold line represents mass signals for species observed when the beam shutter between the Knudsen cell and the cross-beam ionizer was open. The thin line is the background spectrum observed while the shutter was closed. The background pressure in the vacuum chamber was kept below  $1.2 \times 10^{-5}\text{ Pa}$  with an oil-free pumping system. The mass peaks at  $m/e = 52, 53$  and  $54$  indicate the presence of  $\text{CLi}_6$  with natural isotope abundance.





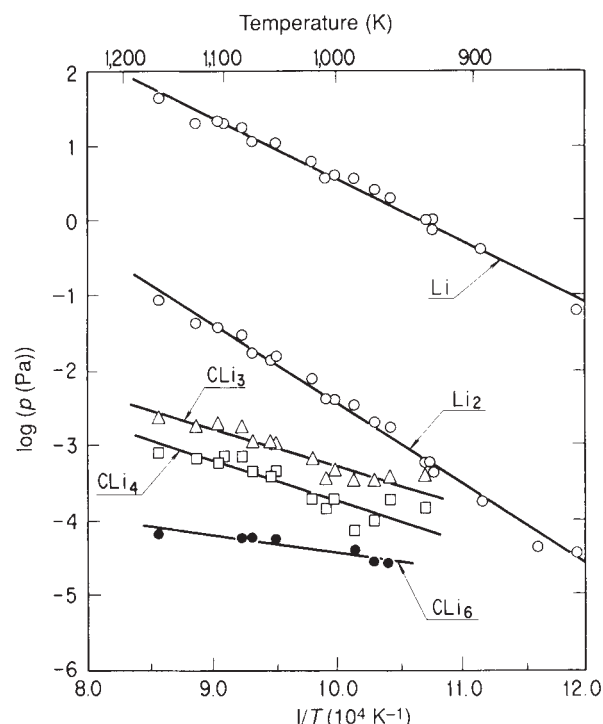


FIG. 3 Equilibrium partial pressures of gaseous species over  $\text{Li}_2\text{C}_2(\text{s})$  as a function of reciprocal temperature.

value  $\Delta H_{298}^\circ = 716.7 \text{ kJ mol}^{-1}$  for the  $\text{C}(\text{s}) \rightarrow \text{C}(\text{g})$  process<sup>22</sup>. The resulting value,  $\Delta H_{298}^\circ = 1,549 \pm 32 \text{ kJ mol}^{-1}$ , is equivalent to the atomization energy of  $D_0^\circ(\text{CLi}_6) = 1,530 \pm 32 \text{ kJ mol}^{-1}$ . The average energy for a C–Li bond in  $\text{CLi}_6$  or contributing bond energy (CBE) derived from the atomization energy becomes  $225 \pm 13 \text{ kJ mol}^{-1}$ . This value is smaller than  $\text{CBE}(\text{CLi}_3) = 306 \pm 15$  and  $\text{CBE}(\text{CLi}_4) = 287 \pm 12 \text{ kJ mol}^{-1}$ , but comparable with  $\text{CBE}(\text{Li}_2\text{O}) = 255 \pm 12 \text{ kJ mol}^{-1}$  of hyperlithiated  $\text{Li}_2\text{O}$  (ref. 17). The large dissociation and atomization energies of  $\text{CLi}_6$  determined here emphasize that strong bonding must be present in such hyperlithiated molecules. Theory shows<sup>3</sup> that the extra electrons beyond the usual octet in  $\text{CLi}_6$  contribute to Li–Li bonding. Experimental evidence for metal–metal bonding in hypervalent species has been reported by Schmidbaur *et al.*<sup>12,13</sup> for polyauroalkane cations:  $\text{Au–Au}$  bonding in  $[(\text{C}_6\text{H}_5)_3\text{PAu}]_5\text{C}^+$  and  $[(\text{C}_6\text{H}_5)_3\text{PAu}]_6\text{C}^{2+}$ . To understand the nature of hypervalent bonding, experiments are continuing, focusing on such molecules as  $\text{CLi}_5$ ,  $\text{Li}_n\text{N}$  and  $\text{Li}_n\text{P}$ . □

20. Wu, C. H. *J. chem. Phys.* **65**, 3181–3186 (1976).
21. Mann, J. B. *J. chem. Phys.* **46**, 1646–1651 (1967).
22. Chase, M. W. *et al. J. Phys. Chem. Ref. Data* **14**, (1985).

ACKNOWLEDGEMENTS. I thank P. v. R. Schleyer for unpublished theoretical data and C. H. Wu for valuable discussion.

## Increase in global atmospheric concentrations of mercury inferred from measurements over the Atlantic Ocean

F. Slemr & E. Langer

Fraunhofer Institute for Atmospheric Environment Research, Kreuzebahnstrasse 19, D-8100 Garmisch-Partenkirchen, Germany

ANTHROPOGENIC processes, such as coal burning, waste incineration and ore refining, are believed to contribute to the emission of mercury into the atmosphere<sup>1–5</sup>. The atmospheric concentration of mercury should therefore have increased as a consequence of increases in anthropogenic emissions<sup>3</sup>. Indeed, analyses of dated soil<sup>6</sup>, peat bog<sup>7,8</sup> and lake-sediment records<sup>8–13</sup> indicate that the deposition of atmospheric mercury may have doubled since the beginning of the nineteenth century. But such an increase in atmospheric mercury concentrations has so far not been detected in ice-core records<sup>14,15</sup> (perhaps because of problems with contamination), and is not consistent with most mercury budgets<sup>2,4,5,16,17</sup>, in which natural sources are thought to dominate. Here we present measurements of total gaseous mercury over the Atlantic Ocean for 1977–90, which show that atmospheric concentrations of mercury have in fact increased by  $1.46 \pm 0.17\%$  per year in the Northern Hemisphere, and by  $1.17 \pm 0.16\%$  per year in the Southern Hemisphere. These rates of increase are consistent with the results of the soil, peat bog and lake-sediment analyses, and lead us to suggest that anthropogenic, rather than natural, sources are at present more important in the mercury cycle.

We measured total gaseous mercury (TGM) in air over the Atlantic Ocean during four ship cruises in the years 1977–1980 (refs 18, 19). There was a pronounced interhemispherical difference in latitudinal TGM distribution, with  $1.96 \pm 0.35 \text{ ng m}^{-3}$  in the Northern Hemisphere and  $1.33 \pm 0.21 \text{ ng m}^{-3}$  in the Southern Hemisphere<sup>18,19</sup>. This difference, and other measurements, suggested<sup>19</sup> that the mean residence time of mercury in the atmosphere was  $\sim 1 \text{ yr}$ , and the tropospheric mercury burden was  $\sim 6 \times 10^9 \text{ g}$ . The data also suggested a trend towards increasing TGM concentration, but because the cruises only spanned a short time, the increase was rather uncertain<sup>20</sup>.

To confirm the trend, we repeated the TGM measurements in October and November 1990 during the RS *Polarstern* cruise from Bremerhaven to Punta Arenas. The itinerary, shown in Fig. 1, was similar to those of the cruises in 1977, 1978 and 1980. TGM was sampled on the uppermost deck,  $\sim 20 \text{ m}$  above sea level. The measurements were made with gold- and silver-coated quartz-wool collectors as used during the previous cruises<sup>19</sup>. The collected mercury was determined by a double amalgamation procedure<sup>21</sup> using an atomic absorption spectrometer (AAS), as during the earlier cruises<sup>22</sup>, or by a resonance fluorescence mercury detector<sup>23</sup> (AFS). The reproducibility of the TGM determination<sup>19</sup> was  $\pm 5.8\%$ , and the measurement uncertainty estimated from the error propagation was less than 15% at the lowest TGM concentrations measured. To characterize air masses and detect contamination by ship exhaust gases, we measured carbon monoxide three times per hour using a gas chromatograph equipped with a HgO detector<sup>24</sup>.

The latitudinal distribution of TGM concentration is shown in Fig. 2. TGM distribution shows essentially three levels of  $\sim 2.2 \text{ ng m}^{-3}$  (STP) north of  $19^\circ \text{ N}$ ,  $1.6 \text{ ng m}^{-3}$  (STP) between

Received 5 July; accepted 5 November 1991.

1. Maercker, A. & Theis, M. *Topics curr. Chem.* **138**, 1–61 (1987).
2. Schleyer, P. v. R. *New Horizon of Quantum Chemistry* (eds Löwdin, F.-O. & Pullman, B.) (Reidel, Dordrecht, 1983).
3. Schleyer, P. v. R., Würthwein, E.-U., Kaufmann, E., Clark, T. & Pople, J. A. *J. Am. chem. Soc.* **105**, 5930–5932 (1983).
4. Würthwein, E.-U., Sen, K. D., Pople, J. A. & Schleyer, P. v. R. *Inorg. Chem.* **22**, 496–503 (1983).
5. Reed, A. D. & Weinhold, F. *J. Am. chem. Soc.* **107**, 1919–1921 (1985).
6. Ritchie, J. P. & Bachrach, S. M. *J. Am. chem. Soc.* **109**, 5909–5916 (1987).
7. Xie, Y. & Schaefer, F. III *Chem. Phys. Lett.* **179**, 563–567 (1991).
8. Grimley, R. T. *The Characterization of High Temperature Vapors* (ed Magrave, J. L.) (Wiley, New York, 1967).
9. Jemmis, E. D. *et al. J. Am. chem. Soc.* **104**, 4275–4276 (1982).
10. Chinn, J. W. Jr. & Lagow, R. J. *J. Am. chem. Soc.* **106**, 3694–3695 (1984).
11. Wu, C. H. & Ihle, H. R. *Chem. Phys. Lett.* **61**, 54–56 (1979).
12. Scherbaum, F., Grohmann, A., Huber, B., Krüger, C. & Schmidbaur, H. *Angew. Chem. Int. Ed. Engl.* **27**, 1544–1546 (1988).
13. Scherbaum, F., Grohmann, A., Müller, G. & Schmidbaur, H. *Angew. Chem. Int. Ed. Engl.* **28**, 463–465 (1989).
14. Grohmann, A., Riede, J. & Schmidbaur, H. *Nature* **345**, 140–142 (1990).
15. Kudo, H., Wu, C. H. & Ihle, H. R. *J. nucl. Mater.* **78**, 380–389 (1978).
16. Wu, C. H., Kudo, H. & Ihle, H. R. *J. chem. Phys.* **70**, 1815–1820 (1979).
17. Wu, C. H. *Chem. Phys. Lett.* **139**, 357–359 (1987).
18. Kudo, H. & Wu, C. H. *Chem. Express* **5**, 633–636 (1990).
19. Kudo, H. *Chem. Lett.* **1989**, 1611–1614.

TABLE 1 Summary of the measurements of total gaseous mercury (TGM) over the Atlantic Ocean

| Cruise                      | Latitude    | TGM concentration (ng Hg m <sup>-3</sup> ) |       |       | No. of samples | VK* (%) |
|-----------------------------|-------------|--------------------------------------------|-------|-------|----------------|---------|
|                             |             | range                                      | mean  | s.d.  |                |         |
| Northern Hemisphere†        |             |                                            |       |       |                |         |
| October 1977‡               | 11–33° N    | 1.0–2.6                                    | 1.763 | 0.362 | 62             | 14.9    |
| Nov./Dec. 1978              | 6–51° N     | 1.42–2.70                                  | 1.849 | 0.306 | 89             | 15.4    |
| Jan./Feb. 1979              | 21–53° N    | 1.63–3.06                                  | 2.169 | 0.382 | 52             | 16.6    |
| Oct./Nov. 1980              | 12–54° N    | 1.41–3.41                                  | 2.085 | 0.351 | 100            | 15.8    |
| Oct./Nov. 1990              | 7–54° N     | 1.41–3.41                                  | 2.247 | 0.409 | 117            | 17.3    |
| Northern Hemisphere average |             |                                            | 2.023 | 0.393 | 420            |         |
| Southern Hemisphere§        |             |                                            |       |       |                |         |
| October 1977‡               | 32° S–11° N | 0.8–1.7                                    | 1.187 | 0.249 | 64             | 15.3    |
| Nov./Dec. 1978              | 18° N–3° N  | 0.86–1.85                                  | 1.350 | 0.207 | 63             | 14.2    |
| Jan./Feb. 1979              | 2° S–4° N   | 1.07–2.09                                  | 1.259 | 0.216 | 37             | 16.2    |
| Oct./Nov. 1980              | 34° S–11° N | 1.10–1.89                                  | 1.453 | 0.157 | 82             | 9.1     |
| Oct./Nov. 1990              | 48° S–7° N  | 0.86–2.44                                  | 1.497 | 0.295 | 158            | 18.8    |
| Southern Hemisphere average |             |                                            | 1.395 | 0.270 | 404            |         |

Measurements made during cruises in 1977–1980 and during the *Polarstern* cruise in October and November 1990.

\* Variation coefficient after correction for the reproducibility of the analytical method;  $VK^2 = VK^2(\text{measured}) - VK^2(\text{analytical})$ , where  $VK(\text{analytical}) = 14.5\%$  during the *Walther Herwig* cruise in 1977 and 5.8% afterwards.

† North of ITCZ.

‡ Corrected for systematic error of the analytical method used during the *Walther Herwig* cruise in 1977.

§ South of ITCZ.

32° S and 19° N, and 1.2 ng m<sup>-3</sup> (STP) south of 32° S. North of 40° N, mercury concentrations in excess of 3 ng m<sup>-3</sup> (STP) coincided with a high CO mixing ratio. A few elevated TGM concentrations were measured at 11° N and at 10° S; at 11° N, but not at 10° S, the high TGM values were accompanied by peak CO values. During the rest of the cruise the low CO mixing ratios of ~65 parts per 10<sup>9</sup> by volume (p.p.b.v.) in the Northern Hemisphere and 45 p.p.b.v. in the Southern Hemisphere indicate that background air masses were sampled. Change of wind direction at 7° N, and climatological data<sup>25</sup>, suggested that the intertropical convergence zone (ITCZ) separating the northern from the southern hemispheric air masses was crossed at ~7° N. Unlike on the former cruises, the passage of ITCZ from the Northern to the Southern Hemisphere was not accompanied by a sudden drop in TGM concentration<sup>18,19</sup>, but it was apparent in a small drop in the CO mixing ratio at 7° N.

The TGM measurements in 1990 are summarized in Table 1 together with the data from previous cruises. Neglecting the high TGM values from 20, 21 and 22 October, the average TGM concentrations over the North and South Atlantic Oceans during October and November 1990 were  $2.25 \pm 0.41$  ( $n = 117$ ) and  $1.50 \pm 0.30$  ( $n = 158$ ) ng Hg m<sup>-3</sup> (STP), respectively. A Mann-Kendall test<sup>26</sup> applied to the mean TGM values of all cruises indicates an increasing trend in both hemispheres at 95% significance level. Linear regression of all individual data points yields a rate of increase of  $1.46 \pm 0.17\% \text{ yr}^{-1}$  ( $n = 420$ ) in the Northern and  $1.17 \pm 0.16\% \text{ yr}^{-1}$  ( $n = 404$ ) in the Southern Hemisphere, both at more than 99.9% significance levels. The trends in the Northern and the Southern Hemispheres differ from each other at the 99% significance level. The observed trends cannot be explained by experimental artefacts, as the analytical procedures have not changed essentially since 1988. In fact, more than half of the analyses were made with the collectors and one third with the AAS instrument used during the 1978–1980 cruises. The AFS instrument, introduced in 1990, was found to yield the same results as the original AAS instrument.

The trend is statistically well supported, but whether the effects is global may be questioned on grounds of its representativeness in space and time. Similar TGM concentrations and the existence of the interhemispherical difference have been reported<sup>5</sup> for the Pacific Ocean. This agreement is not surprising, in view of the well-accepted TGM residence time<sup>2,5,19</sup> in the atmosphere of ~1 year. The TGM residence time is about four times as long as the intrahemispherical mixing time<sup>27</sup> of ~3 months and, consequently, both hemispheres are likely to be well mixed with

respect to the TGM concentrations. All cruises were made in the same season, and three of them were made in October and November. Nevertheless, fluctuations of the seasonal variation such as reported by Brosset<sup>28</sup> may still bias the trend. To assess this bias, we monitored TGM concentrations on top of the Wank mountain (1,780 m above sea level) in the Bavarian Alps. More than 500 individual TGM measurements made between March 1990 and August 1991 showed a substantial variability, most of which could be attributed to the advection of polluted air masses. In all seasons we observed background values of ~2 ng m<sup>-3</sup> in

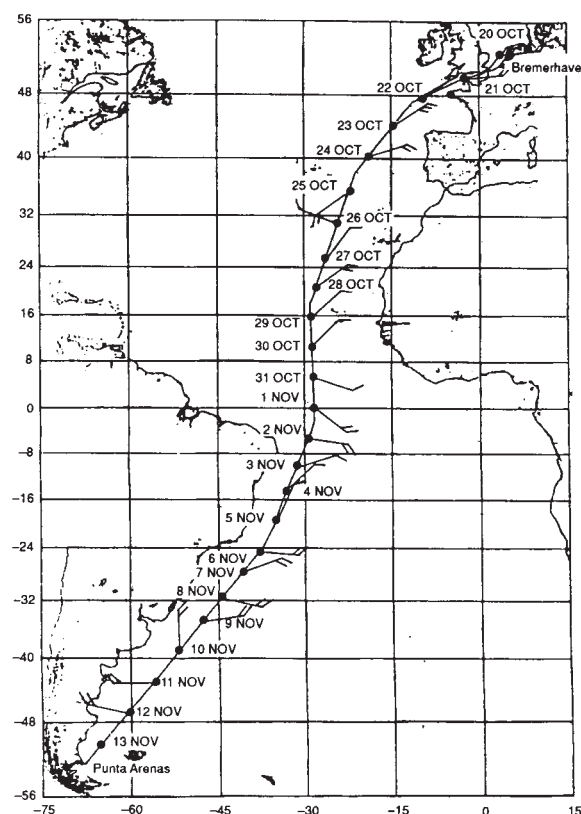
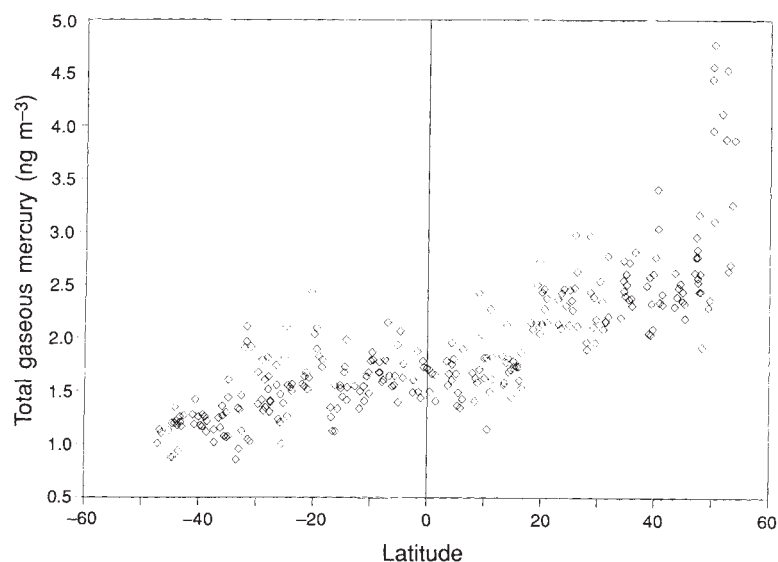


FIG. 1 Itinerary of the *Polarstern* cruise from Bremerhaven to Punta Arenas in October and November 1990. The arrows indicate the wind direction and the number of flags indicate with wind speed in units of  $5 \text{ m s}^{-1}$ .



FIG. 2 Total gaseous mercury concentrations as a function of latitude. The concentrations are in  $\text{ng Hg m}^{-3}$  (STP).



air masses of marine origin with low levels of anthropogenic pollutants. The random appearance of these background values did not suggest systematic seasonal variations of background TGM concentration at middle northern latitudes. This finding is again consistent with long TGM residence time in the atmosphere which makes large seasonal variation unlikely. Consequently, we conclude that the observed trend is not biased by seasonal variations and is global in character. Additionally, TGM values of  $\sim 1 \text{ ng m}^{-3}$  at Los Altos near San Francisco and over the Pacific Ocean measured by a similar technique<sup>29</sup> in 1965–1967 indicate that the TGM concentration may have been increasing for more than 25 years.

The global increase in TGM concentration may be the result of decreasing sink strength or increasing emission rates, or a combination of both. More than 92% of background TGM consists of elemental mercury<sup>19</sup>. Its high vapour pressure and low solubility prevent effective removal by attachment to aerosols or by wet deposition. It is assumed that elemental mercury has to be oxidized to a compound more suitable for deposition, but the processes involved are not well known<sup>30</sup>. Some laboratory studies<sup>31–33</sup> suggest that reactions with ozone or hydrogen peroxide initiate the deposition process. Observations<sup>34</sup> and model calculations<sup>35–37</sup> indicate that tropospheric ozone concentrations are increasing in the Northern Hemisphere and are stable or decreasing in the Southern Hemisphere. These trends would imply a decrease of the TGM concentration in the Northern Hemisphere and a stable TGM concentration in the Southern Hemisphere. The observed TGM trends show an opposite pattern, indicating that increasing emissions are probably the dominant cause of the trend.

The increasing emissions are probably due to increasing anthropogenic activities. About 65% of anthropogenic mercury emission to the atmosphere is due to coal burning, and another 25% is due to waste incineration<sup>1</sup>. Worldwide coal production increased at an annual rate of 2.1% between 1965 and 1987, and  $\sim 90\%$  of coal was burned in the Northern Hemisphere<sup>38</sup>. Taking into account that flue gas cleaning in the developed countries may substantially reduce the mercury emission<sup>30</sup> and a contribution due to the coal burning, the observed TGM trends in both hemispheres are in rough agreement with the upward trend of coal consumption. On the other side, estimates of natural mercury emissions are highly uncertain<sup>39</sup> and the more recent ones tend to lower values<sup>2,4,5,16,17,30,39</sup>. Lindqvist *et al.*<sup>30</sup> estimate that natural emissions may represent only 40% of total emissions. Even this contribution may be too high in view of the anthropogenic emission of  $4.5 \times 10^9 \text{ g Hg yr}^{-1}$  estimated by Lindqvist *et al.*<sup>30</sup> and the total flux of  $6 \times 10^9 \text{ g Hg yr}^{-1}$  derived from our measurements<sup>19</sup>. This impression is further supported by

recent lidar measurements of mercury in air over geothermally active region of Iceland<sup>40</sup> which indicate that emission from geothermal regions may be overestimated. Our measurements have also provided no indication of the proposed<sup>41</sup> mercury emission from oceanic upwelling regions.

The observed rate of increase of TGM concentration in air and its attribution to the increase of anthropogenic emission are consistent with the analyses of peat bog, soil and sediment horizons in the United States, Canada, Sweden, Finland and Germany<sup>6–13</sup> suggesting that since the beginning of the nineteenth century, the global deposition of atmospheric mercury may have increased by a factor ranging from 0.5 to 5.1. If substantiated by further measurements, so large a global TGM increase will be of concern because of mercury enrichment in the aquatic food chain and its conversion to highly toxic alkyl compounds<sup>30</sup>. □

Received 8 October; accepted 19 November 1991.

- Nriagu, J. O. & Pacyna, J. M. *Nature* **333**, 134–139 (1988).
- Lindqvist, O. & Rodhe, H. *Tellus* **B37**, 136–159 (1985).
- Watson, W. D. Jr in *The Biogeochemistry of Mercury in the Environment*, 41 (Elsevier, Amsterdam, 1979).
- Lindberg, S. E. in *Lead, Mercury, Cadmium and Arsenic in The Environment*, 89 (Wiley, Chichester, 1987).
- Fitzgerald, W. F. in *The Role of Air–Sea Exchange in Geochemical Cycling*, NATO ASI Ser. **C185** 363 (1986).
- Kaiser, G., Tölg, G. & Schlichting, E. in *Sonderreihe Umwelttagung* (University of Hohenheim, 1978).
- Pheiffer Madson, P. *Nature* **293**, 127–130 (1981).
- Norton, S. A., Dillon, P. J., Douglas Evans, R., Mierle, G. & Kahl, J. S. in *Acidic Precipitation*, Vol. 3, 73 (Springer, Berlin, 1990).
- Heit, M., Tan, Y., Klusek, C. & Burke, J. C. *Water Air Soil Pollut.* **15**, 441–464 (1981).
- Tolonen, K. & Jaakkola, T. *Ann. Bot. Fennici* **20**, 57–78 (1983).
- Johansson, K. *Verh. Internat. Verein. Limnol.* **22**, 2359–2363 (1985).
- Douglas Evans, R. *Arch. Envir. Contam. Toxicol.* **15**, 505–512 (1986).
- Rada, R. G., Wiener, J. G., Winfrey, M. R. & Powell, D. E. *Arch. Envir. Contam. Toxicol.* **18**, 175–181 (1989).
- Wolff, E. W. & Peel, D. A. *Nature* **313**, 535–540 (1985).
- Peel, D. A. in *The Environmental Record in Glaciers and Ice Sheets*, 207 (Wiley, Chichester, 1989).
- Andren, A. W. & Nriagu, J. O. in *The Biogeochemistry of Mercury in the Environment*, 1 (Elsevier–North-Holland, Amsterdam, 1979).
- Kothny, E. L. in *Trace Elements in the Environment*, 48 (American Chemical Society, Washington, 1973).
- Slemr, F., Seiler, W. & Schuster, G. *J. geophys. Res.* **86**, 1159–1166 (1981).
- Slemr, F., Seiler, W. & Schuster, G. *J. Atmos. Chem.* **3**, 407–434 (1985).
- Slemr, F., Seiler, W. & Schuster, G. in *Atmosphärische Spurenstoffe*, 197 (Verlag Chemie, Weinheim, 1986).
- Slemr, F., Seiler, W. & Schuster, G. *Ber. Bunsenges. Phys. Chem.* **82**, 1142–1146 (1978).
- Slemr, F., Seiler, W., Eberling, C. & Roggendorf, P. *Analyt. chim. Acta* **110**, 35–47 (1979).
- Thompson, K. C. & Godden, R. G. *Analyst* **103**, 544 (1975).
- Seiler, W., Giehl, H. & Roggendorf, P. *Atmos. Technol.* **12**, 40 (1980).
- Marine Climatological Summary for the Atlantic Ocean 1971–1980*, No. 12 (Seewetteramt, Deutscher Wetterdienst, Hamburg, 1987).
- Gilbert, R. O. *Statistical Methods for Environmental Pollution Monitoring* (Van Nostrand Reinhold, New York, 1987).
- Warneck, P. *Chemistry of the Natural Atmosphere*, 158 (Academic, New York, 1988).
- Brosset, C. *Water Air Soil Pollut.* **17**, 37 (1982).
- Williston, S. H. *J. geophys. Res.* **73**, 7051 (1968).
- Lindqvist, O. *et al. Water Air Soil Pollut.* **55**, 1 (1991).
- Pjankov, V. A. *Zhur. Obshestven. Khim.* **19**, 224 (1949).

32. Iverfeldt, A. & Lindqvist, O. *Atmos. Environ.* **20**, 1567 (1986).
33. Kobayashi, T. *J. Japan Soc. Air Pollut.* **22**, 230 (1987).
34. Bojkov, R. D. in *Tropospheric Ozone*, 83 (Reidel, Dordrecht, 1988).
35. Crutzen, P. J. & Brühl, C. in *The Environment Record in Glaciers and Ice Sheets*, 249 (Wiley, Chichester, 1989).
36. Liu, S. C. in *Tropospheric Ozone*, 303 (Reidel, Dordrecht, 1988).
37. Thompson, A. M., Huntley, M. A. & Stewart, R. W. *J. geophys. Res.* **95**, 9829 (1990).
38. *Reserven, Ressourcen und Verfügbarkeit von Energierohstoffen* (Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, 1989).
39. Nriagu, J. O. *Nature* **338**, 47–49 (1989).
40. Edner, H. et al. *J. geophys. Res.* **96**, 2977–2986 (1991).
41. Fitzgerald, W. F., Gill, G. A. & Kim, J. P. *Science* **224**, 597–599 (1984).

ACKNOWLEDGEMENTS. We thank the Alfred-Wegener-Institute and the crew of RS *Polarstern*. We especially thank E. Scheel and H. Cassebaum who helped us to prepare the CO measurements, W. Seiler for comments on the manuscript, and M. Schulz-Baldes and O. Schrems who helped us with the measurements on RS *Polarstern*. We acknowledge the Swedish Environmental Protection Boards invitation to F.S. to attend the workshop on mercury in the environment in Gavle, 1990. This work was supported by the Deutsche Forschungsgemeinschaft.

## Large-scale mantle convection and the history of subduction

Mark A. Richards\* & David C. Engebretson†

\* Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

† Department of Geology, Western Washington University, Bellingham 98225, USA

A LARGE-SCALE pattern of mantle convection is evident in the shape of the Earth's gravity field<sup>1</sup>, lateral variations in seismic velocity<sup>2,3</sup>, the distribution of hotspots<sup>4</sup>, and long-wavelength bathymetric anomalies<sup>5</sup>. Paradoxically, these signals of mantle convection bear little resemblance to present-day plate motions<sup>1,4</sup>, although they seem to be related to past subduction-zone configurations<sup>6</sup>. Here we quantify this relationship by correlating a time-integrated history of subduction, calculated in the hotspot reference frame for the past 180 Myr, with regions of the deep mantle inferred from other geophysical observations to be colder than average. Our results support suggestions<sup>6–9</sup> that a direct relationship exists between large-scale thermal heterogeneity in the lower mantle and subduction during the Cenozoic and Mesozoic.

The observed geoid<sup>10</sup>, corrected for the Earth's hydrostatic figure<sup>11</sup> and for modelled effects of seismically active slabs in the upper mantle<sup>8</sup>, is shown in Fig. 1a. Recent geodynamical models<sup>3,12–15</sup> explain much of the low-harmonic-degree ( $L = 2, 3$ ) power in the observed geoid in terms of seismically inferred density anomalies. These models rely primarily on the observation that broad regions of low seismic velocity (low density) underlie the large residual geoid highs, and *vice versa*, with the geoid being dominated by dynamically supported topography<sup>16,17</sup>. Figure 1b shows the  $L = 2$  and 3 components of Dziewonski's lower-mantle P-wave model<sup>2</sup>, radially averaged, for comparison with the geoid. That this pattern indeed marks regions of high and low density and, presumably, low and high temperatures is reinforced by the observation that most volcanic hotspots occur in areas of inferred high temperatures<sup>4,9</sup> (Fig. 1c). A degree-2 pattern of dynamic topography with amplitude ~250 m, inferred from anomalous sea-floor elevations, also matches the pattern predicted by the geoid models<sup>5</sup>.

Anderson<sup>18</sup> suggested that this pattern results from insulation of the mantle beneath the pre-Mesozoic Pangean continent. Recognizing that long-term insulation (stagnation) of the upper mantle is unlikely, Chase and Sprowl<sup>6</sup> showed that the circum-Pacific band of geoid lows corresponded to the pattern of Mesozoic subduction at ~125 Myr. They suggested that lower-mantle convection and plate tectonics are coupled on a timescale of at least 100 Myr, because of either layered convection or weak coupling between plate motions and the deep mantle. Davies<sup>7</sup> showed with simple convection models that mantle heterogeneity structure could lag behind rapidly changing plate boundary configurations, and that the geoid could plausibly

represent a thermal structure developed in the Mesozoic or earlier. We showed previously<sup>9</sup> that the negative buoyancy flux of cold subducted slabs during the Cenozoic and Mesozoic was, in principle, sufficient to explain the degree-2 seismic velocity structure of the lower mantle (and geoid). Thus a time-integrated accounting for subducted buoyancy over the past ~100–200 Myr might explain the observed large-scale structure of convection (assuming, implicitly, that mid-ocean ridges are passive structures and that mantle heterogeneity is primarily the result of subduction).

To test this hypothesis quantitatively, we have reconstructed the amount of plate area subducted and its location in the hotspot reference frame throughout the Cenozoic and Mesozoic. This is increasingly difficult as we go back in time, and reconstructions are poorly constrained before ~150 Myr. Details of our methods are given in previous work<sup>19</sup> on Pacific basin plate motions since 180 Myr, and we have extended this work to include a global mapping of subduction using an updated compilation of finite plate rotations<sup>20</sup>. The results are shown in Fig. 2a–f. Convergent boundaries are shown in their time-averaged position (in the hotspot reference frame) during 30-Myr intervals, with the relative amounts of plate area subducted along these convergent boundary segments represented by the areas of the plotted squares. (Boundaries are projected 700 km in the direction of the overriding plate.) In each 30-Myr window shown, the subducted plate area is calculated during 5-Myr intervals according to stage poles of relative plate motion<sup>19</sup>. The plate boundaries shown are approximate, but their detailed shapes are unimportant; the amount of plate subducted along a convergent boundary is only a weak function of undulations in the boundary. In calculating the longest-wavelength components (~10,000–20,000 km for  $L = 2, 3$ ) of the spatial distribution of subducted plate area, location inaccuracies as large as several thousand kilometres are of little consequence.

Figure 1d shows the  $L = 2, 3$  component of subducted plate area, integrated from 0–120 Myr. This representation was obtained by spherical harmonic expansion of the subducted plate area patches shown in Fig. 2. We make no effort to guess the trajectories of old slabs after they penetrate through the upper mantle. The field shown in Fig. 1d is, effectively, a low-pass-filtered map of ancient subduction zone locations weighted by subduction rates. (If we were constructing a dynamical model, which we are not, this would amount to assuming that horizontal motions in the deep mantle are much less than ~10,000 km over timescales of ~100 Myr; that is, much less than 10 cm yr<sup>-1</sup>. Relative motions among hotspots, which are presumably rooted in the deep mantle, clearly meet this criterion.)

The low-degree distribution of cold subducted slabs (blue areas of Fig. 1d) corresponds well with the inferred large-scale pattern of cold downwellings in the mantle (Fig. 1a–c). Better insight into this observation is gained from examining formal correlations and their variation through geological time.

Figure 3a shows the  $L = 1–4$  correlations<sup>21</sup> between the radially averaged lower-mantle seismic velocity structure of Dziewonski<sup>2</sup> and subducted slab flux as a function of time. Degree-1 correlations are included because, unlike the geoid, both the slab flux and seismic structure contain a degree-1 term. Each set of points represents a 30-Myr time interval; for example, the points plotted at 45 Myr represent the correlation between the 30–60-Myr slab flux and seismic velocity. Positive correlations signify that slabs correspond to high-velocity anomalies. For 0–30 Myr the correlation is positive but not statistically significant. Correlations for degrees 1–3 are much stronger for the time intervals 30–60, 60–90 and 90–120 Myr. The poorer correlations at 150 and 180 Myr may not be meaningful, owing to uncertainties in the older reconstructions. Correlations of degree 4 (and higher) are never significant, as expected from the lack of agreement among seismic velocity models at higher



degree as well as from their relatively poor correlation with the geoid<sup>13</sup>.

Figure 3*b* shows similar correlations, except that the time axis refers to total integration time: points plotted at 90 Myr represent all subducted slab flux from 0 to 90 Myr. Correlations increase as we go back in time and stay high for  $L = 1-3$  before 60 Myr, with all degree-2 correlations statistically significant at >95% confidence level. The correlations of Fig. 3*a* and *b* are similar if other lower-mantle seismic velocity models are used<sup>9,23-25</sup>, although the details vary.

Figure 3*c* shows the time-integrated correlations between subducted slab flux and the residual geoid for degrees 2-4. Positive correlations indicate that geoid lows (lower-mantle density highs) correspond to the distribution of subducted slabs. Again, the nonhydrostatic geoid has been corrected for the signal over subduction zones active at present<sup>8</sup>, so the residual geoid and subducted slab flux are uncorrelated during the time interval 0-30 Myr. We feel that this corrected geoid is a better indicator of deep-mantle heterogeneity, even though the results are not

qualitatively different if this correction is omitted. Like the correlations with seismic velocity, the  $L = 2-3$  correlations between slab flux and the geoid increase as we go back in time, with the best correlations for the time-integrated flux occurring before 90 Myr.

These results show that the spatial distribution of plate area subducted since ~90-180 Myr, calculated in the hotspot reference frame, matches the large-scale pattern of anomalously cold regions in the deep mantle inferred from seismology and the geoid. It is also easily shown<sup>9</sup> that the thermal load of slabs subducted is sufficient to explain the amplitude of the  $L = 2, 3$  geoid anomalies and of the low-degree temperature variations of ~50 °C inferred from seismic tomography. Thus one could, for example, construct a model for the present-day geoid (and seismic heterogeneity) by assuming an isothermal mantle at 120 Myr and distributing the thermal load of the slabs of Fig. 2*a-d* through the depth of the lower mantle. Such a model is oversimplified, but it explains as much or more of the geoid signal as models based upon seismic tomography and simul-

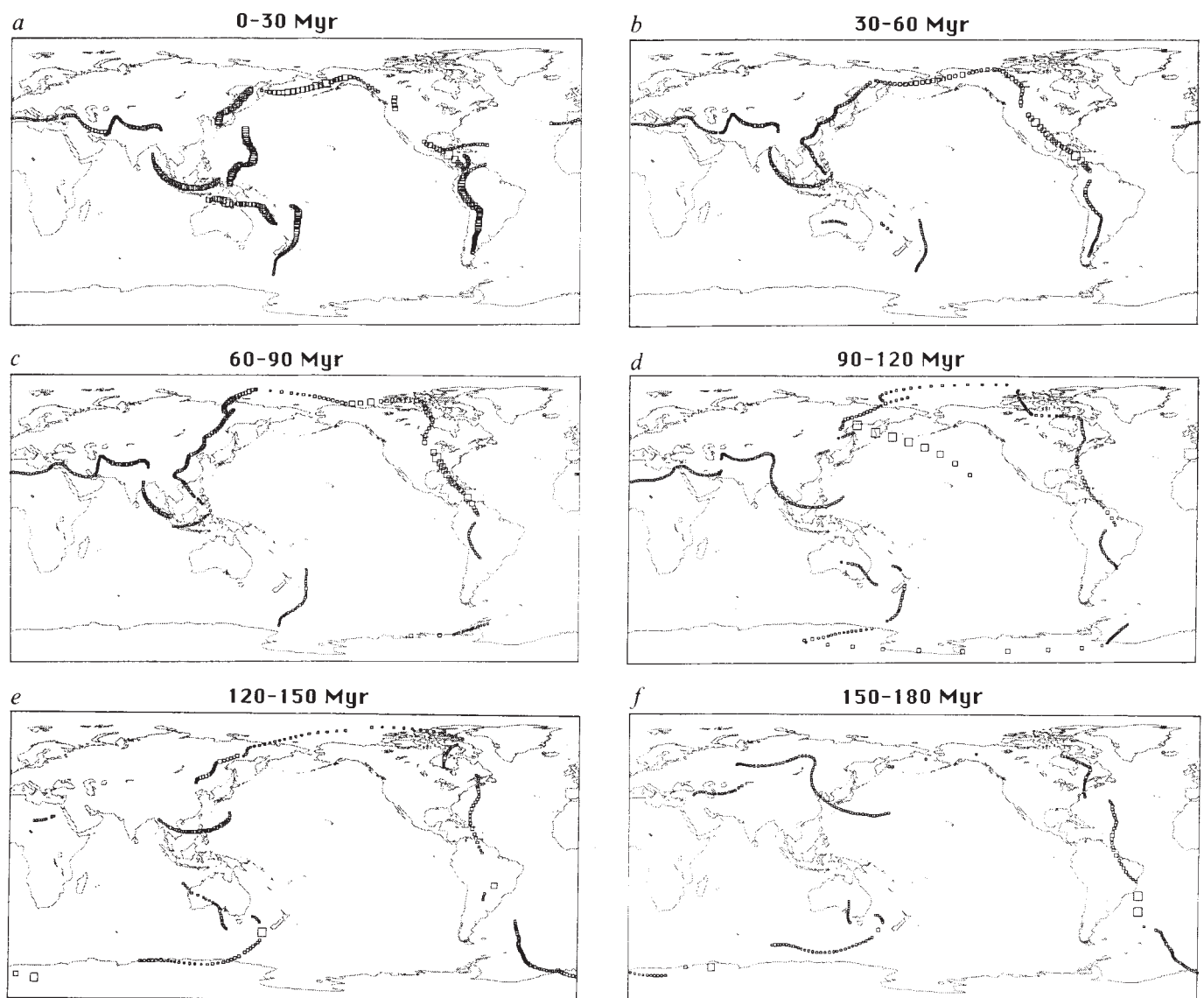


FIG. 2 Subduction histories for 30-Myr time intervals from 0-180 Myr, with continental outlines in present-day positions. Plotted squares indicate average locations of subduction zones during each time interval. The relative amount of plate area subducted along each convergent boundary segment during the time interval is proportional to the area of the square, with each plot normalized to the maximum within that time interval. Within the 30-Myr intervals, the subducted plate area is calculated for each segment by

summing over six 5-Myr subintervals, using the appropriate relative motion stage pole for each subinterval. Average subduction zone positions are projected horizontally 700 km in the direction of the overriding plate. (Otherwise, no other assumption is made concerning the ultimate fate of the slab.) The largest square in each plot represents: *a*,  $0.68 \times 10^6$ ; *b*,  $2.15 \times 10^6$ ; *c*,  $1.91 \times 10^6$ ; *d*,  $4.54 \times 10^6$ ; *e*,  $2.93 \times 10^6$ ; *f*,  $4.21 \times 10^6$  km<sup>2</sup>.

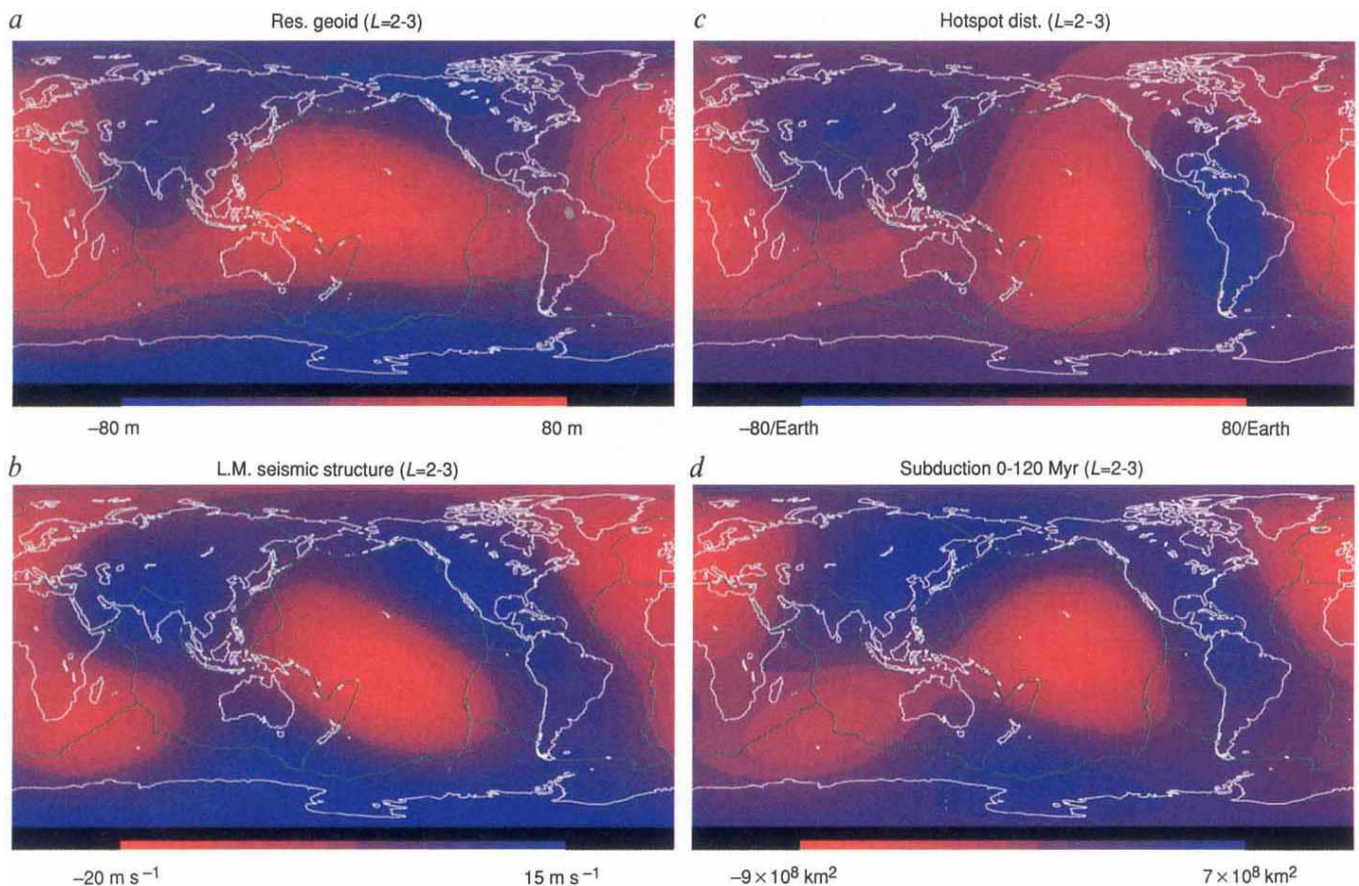


FIG. 1 *a*, Observed geoid<sup>10</sup>, corrected for hydrostatic figure<sup>11</sup> and upper-mantle slab signal<sup>8</sup>. *b*, Radially averaged lower-mantle seismic velocity model

of Dziewonski<sup>2</sup>. *c*, Hotspot distribution<sup>9</sup>. *d*, Time-integrated flux of subducted slabs since 120 Myr, calculated in the hotspot reference frame.

taneously provides a clear dynamical link between plate tectonics and mantle heterogeneity structure.

Numerical models of mantle convection suggest several interpretations of the apparent time delay between changes in the pattern of subduction and changes in large-scale mantle heterogeneity. (1) A high-viscosity lower mantle would increase the transit time of subducted slabs through the mantle<sup>27</sup>, and it might also explain the very existence of the hotspot reference frame in which our plate reconstructions were made<sup>28-30</sup>. (2) The pattern of convection may be stabilized at long wavelengths by supercontinent aggregation, because continental lithosphere is not readily subducted<sup>31,32</sup>. More generally, there can be a long time lag between rapid changes in plate motions and changes in large-scale mantle heterogeneity<sup>7</sup>. (3) Mantle convection may be layered, with the lower mantle thermally coupled to subduction on timescales  $>100$  Myr (ref. 6).

With regard to these mechanisms, we add that rates of global subduction and spreading during the period  $\sim 90$ – $140$  Myr were almost double those of the late Cenozoic (0–30 Myr)<sup>20</sup> (see scaling factors in Fig. 2). The disproportionate amount of slab heterogeneity introduced into the mantle during the Cretaceous may explain why the correlations of Fig. 3*b* and *c* go up and remain high as we go back into the Mesozoic. Thus our results suggest that the correspondence between Mesozoic subduction zones and present-day mantle structure may result as much from the variability or episodicity of plate motions as from any inherent dynamical time lag.

For the same reason, the large-scale heterogeneity structure of the mantle could have changed substantially since  $\sim 100$  Myr,

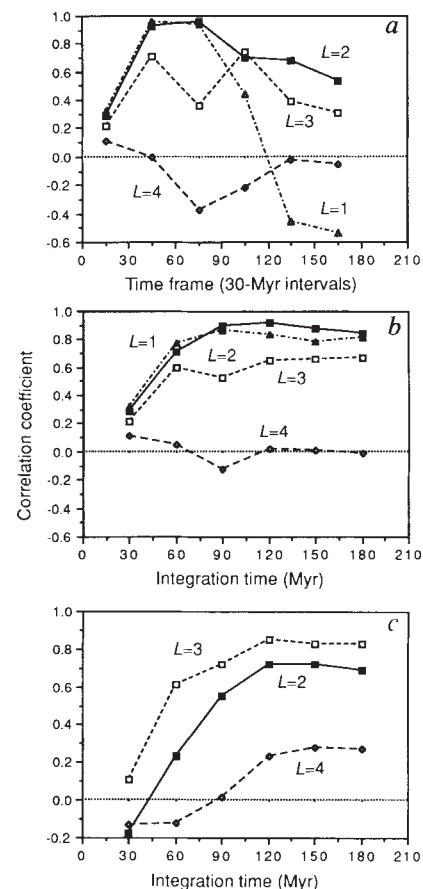


FIG. 3 Correlations<sup>21</sup> between subduction history and (*a, b*) lower-mantle seismic velocity and (*c*) residual geoid. Curves are for correlations at harmonic degrees  $L=1$  ( $\Delta$ ),  $L=2$  ( $\blacksquare$ ),  $L=3$  ( $\diamond$ ) and  $L=4$  ( $\times$ ). See text for further explanation.



along with the geoid and dynamic topography. Thus the answer to the paradox posed by Gurnis<sup>26</sup>, regarding the lack of observed continental flooding in response to Phanerozoic continental motions in a 'static' geoid-topography field, may be simply that neither the long-wavelength geoid nor associated dynamic topography have remained static.

Numerical mantle convection models are now approaching sufficient resolution to address such time-dependent problems in global geodynamics, and our results show how detailed knowledge of the history of plate motions can be used to constrain these models. For this purpose, estimates for the ages of subducted plates would be desirable so that their thermal loads could be estimated more accurately. □

Received 10 July; accepted 3 December 1991.

1. Chase, C. G. *Nature* **282**, 464–468 (1979).
2. Dziewonski, A. M. *J. geophys. Res.* **89**, 5929–5952 (1984).
3. Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M. *Nature* **313**, 541–515 (1985).
4. Crough, S. T. & Jurdy, D. M. *Earth planet. Sci. Lett.* **48**, 15–22 (1980).
5. Cazenave, A. & Lago, B. *Nature* **340**, 54–57 (1989).
6. Chase, C. G. & Sprowl, D. R. *Earth planet. Sci. Lett.* **62**, 314–320 (1983).
7. Davies, G. F. *Earth planet. Sci. Lett.* **69**, 187–194 (1984).

8. Hager, B. H. *J. geophys. Res.* **89**, 6003–6015 (1984).
9. Richards, M. A., Hager, B. H. & Sleep, N. H. *J. geophys. Res.* **93**, 7690–7708.
10. Lerch, F. J., Klosko, S. M. & Patch, G. B. *NASA Tech. Memo TM84986* (1983).
11. Nakiboglu, S. M. *Phys. Earth planet. Inter.* **28**, 302–311 (1982).
12. Forte, A. M. & Peltier, W. R. *J. geophys. Res.* **92**, 3645–3679 (1987).
13. Richards, M. A. & Hager, B. H. in *The Physics of Planets* (ed. S. K. Runcorn) 247–272 (Wiley, Chichester, 1988).
14. Ricard, Y., Froidevaux, C. & Fleitout, L. *Geophys. J. Int.* **93**, 477–484 (1988).
15. Hager, B. H. & Richards, M. A. *Phil. Trans. R. Soc. Lond.* **A328**, 309–327 (1989).
16. Richards, M. A. & Hager, B. H. *J. geophys. Res.* **89**, 5987–6002 (1984).
17. Ricard, Y., Fleitout, L. & Froidevaux, C. *Ann. Geophys.* **2**, 267–286 (1984).
18. Anderson, D. L. *Nature* **297**, 391–393 (1982).
19. Engebretson, D. C., Cox, A. & Gordon, R. G. *Geol. Soc. Am. Spec. Pap.* **206** (1985).
20. Engebretson, D. C., Kelly, K. P., Cashman, H. J. & Richards, M. A. *Eos* **71**, 1575 (1990).
21. Ekhhardt, D. H. *Math. Geol.* **16**, 155–171 (1984).
22. Hager, B. H. & Clayton, R. W. in *Mantle Convection* (ed. W. R. Peltier) 657–763 (Gordon and Breach, New York, 1988).
23. Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y. *Phys. Earth planet. Int.* **59**, 294–328 (1990).
24. Tanimoto, T. *Geophys. J. Int.* **100**, 327–336 (1990).
25. Vasco, D. W., Pulliam, R. J. & Johnson, L. R. *J. geophys. Res.* (in the press).
26. Gurnis, M. *Nature* **344**, 754–756 (1990).
27. Bunge, H.-P. & Richards, M. A. *Eos* **71**, 1625 (1990).
28. Gurnis, M. & G. F. Davies, *Geophys. J. R. astr. Soc.* **85**, 523–541 (1986).
29. Duncan, R. A. & Richards, M. A. *Rev. Geophys.* **29**, 31–50 (1991).
30. Richards, M. A. in *Glacial Isostasy, Sea-Level and Mantle Rheology* (eds Sabadini, R. et al.) 571–587 (Kluwer, Amsterdam, 1991).
31. Gurnis, M. & Zhong, S. *Geophys. Res. Lett.* **18**, 581–584 (1991).
32. Gurnis, M. *Nature* **322**, 695–699 (1988).

ACKNOWLEDGEMENTS. This work was supported by the NSF. We thank C. Chase for reviewing the manuscript.

## Why do female adders copulate so frequently?

Thomas Madsen<sup>\*†</sup>, Richard Shine<sup>†‡</sup>, Jon Loman<sup>\*</sup> & Thomas Håkansson<sup>§</sup>

<sup>\*</sup> Department of Animal Ecology, University of Lund, S 223 62 Lund, Sweden

<sup>†</sup> Zoology Department, The University of Sydney, New South Wales 2006, Australia

<sup>§</sup> Department of Anthropology, University of Kentucky, Lexington, Kentucky 40506, USA

**MALES of most animal species will enhance their reproductive success if they mate often and with many different partners, whereas promiscuous mating is unlikely to increase a female's reproductive success. Why then is multiple copulation by females so common<sup>1–6</sup>? Many theoreticians have suggested that multiple copulations might enhance the viability of a female's offspring, either because of inadequate quantities of sperm from the first mating<sup>1,7</sup>, additional nutrients derived from the seminal fluid<sup>7,8</sup> or some genetic advantage<sup>9–14</sup>. Our field studies on Swedish adders provide the first empirical evidence that multiple copulations, with different partners each time, increase offspring viability. This advantage apparently results from more intense sperm competition in the female's reproductive tract, resulting in a higher proportion of her ova being fertilized by genetically superior males.**

Adders (*Vipera berus*) are small venomous snakes widely distributed through Europe. Like all snakes, they lack male parental care<sup>15</sup>, and female cooperation is required for successful copulation (forcible insemination is impossible<sup>12</sup>). After mating, females store sperm in their reproductive tracts for months before ovulation<sup>12,16,17</sup>. For the past 10 years we have studied a small population of adders (mean number of reproducing adults per annum = 30, range 21–46) living in open grassy meadows in southern Sweden (Smygehuk, 55° 20' N, 13° 22' E). All snakes captured in the study area were individually marked by clipping ventral scales, and all reproductive females were force-fed miniature radiotransmitters each year. The snakes were monitored intensively throughout the three-week mating period each spring. Females were collected in midsummer, and maintained in the laboratory until they gave birth a week or two later. The neonates were measured, weighed, marked and

released, with their mothers, at her site of capture.

Most females mated with more than one male. The average number of copulations per female per season was 3.69 (s.d. = 1.58,  $n = 45$  females), and ranged from one to eight. The number of times that a female copulated was independent of her body length ( $r = 0.18$ ,  $n = 45$ ,  $P = 0.23$ ) or the number of males that we saw courting her during the mating season ( $r = 0.11$ ,  $n = 45$ ,  $P = 0.45$ ). Female adders also tended to mate with a series of different males, rather than repeatedly with the same male. A female adder was courted by an average of seven males per season, and mated at least three times. Hence, the probability that she would have mated more than once with the same male is 38.5% under the null hypothesis of random choice of partners. In practice, the observed proportion of 'repeat matings' was much lower than this figure (16.1% of 124 matings), and the difference between these two proportions is highly significant ( $\chi^2 = 26.21$ , 1 d.f.,  $P < 0.001$ ). This test is conservative, because it does not take into account the fact that females actually mated an average of 3.7 times rather than 3.0, and that some males were much more successful than others in obtaining matings. These factors should have considerably increased the numbers of 'repeat' matings.

Why do female adders mate more than once per season, and usually with a different male each time? The female's reproductive success (litter size) is determined by the amount of energy she allocates to reproduction, so her reproductive output is not enhanced by additional matings: the number of copulations by a female was not significantly correlated with her litter size ( $r = 0.15$ ,  $n = 30$ ,  $P = 0.44$ ) or with her fecundity relative to body size (residual scores from the regression of litter size to maternal snout-vent length:  $r = 0.26$ ,  $n = 30$ ,  $P = 0.17$ ). The nutrient input from semen is insignificant: the number of matings by a female did not affect her mean offspring mass ( $r = 0.21$ ,  $n = 28$ ,  $P = 0.29$ ), her total litter mass ( $r = 0.17$ ,  $n = 27$ ,  $P = 0.38$ ) or her proportional loss in body mass during gestation ( $r = 0.25$ ,  $n = 28$ ,  $P = 0.21$ ). A shortage of sperm to fertilize her ova is also not a problem: studies on captive adders show that a single mating is sufficient for this purpose<sup>18</sup>, and our field data show that almost all ova were fertilized in almost all litters (mean = 3.9% infertile, s.d. = 8.5,  $n = 34$  litters), and that the proportions of unfertilized ova were not affected by the number of times that a female mated ( $r = 0.02$ ,  $n = 34$ ,  $P = 0.90$ ).

Our data suggest, instead, that a female adder benefits from multiple copulations through the genetic quality of her offspring. Genetic factors may be particularly important in small isolated

† To whom all correspondence should be addressed.

populations such as that of the Smygehuk adders, because of the high potential for inbreeding depression<sup>19</sup>. The mating system of adders, whereby a single large male may obtain most of the matings, reduces effective population size and thus amplifies this effect<sup>20</sup>. The Smygehuk adders offer an ideal opportunity to look for an influence of multiple mating on offspring viability, because genetic bottlenecks and small effective population sizes have resulted in low genetic variability and high proportions of stillborn offspring in most litters (mean = 31.6% of offspring stillborn, s.d. = 27.4, range = 0–100%,  $n = 34$  litters).

Multiple matings strongly reduced the proportions of offspring that were dead at birth (correlating proportion of stillborn young against number of matings,  $r = -0.40$ ,  $n = 34$ ,  $P < 0.02$ ; against number of different males mated with,  $r = -0.40$ ,  $n = 34$ ,  $P < 0.02$ ). No other variable that we measured on females or their litters was significantly correlated with the proportion of stillborn young (for example litter size,  $r = 0.04$ ,  $n = 34$ ,  $P = 0.25$ ; mean offspring mass,  $r = 0.24$ ,  $n = 30$ ,  $P = 0.20$ ), and there were no consistent differences in this respect between first and second litters of the same female ( $t$ -test, 8 d.f.,  $t_8 = 0.24$ ,  $P = 0.81$ ), or among individual females (using a one-factor analysis of variance, with identification number of the female as the factor,  $F_{5,6} = 1.18$ ,  $P = 0.42$ ). The number of times that a female mated also did not differ consistently among individual females between years ( $F_{5,6} = 1.24$ ,  $P = 0.39$ ), or between first and second litters ( $t_8 = 0.79$ ,  $P = 0.45$ ). Hence the number of matings (or numbers of different mates, which is highly correlated with number of matings:  $r = 0.90$ ,  $n = 34$ ,  $P < 0.001$ ) seems to be the primary determinant of the proportion of viable offspring produced by a female adder.

More detailed analysis shows that offspring viability depends on a female's choice of mates as well as the number of times she copulates. We looked for a paternal effect on offspring viability by restricting analysis to litters that had been sired by only one or two males, and comparing the proportion of stillborn offspring among fathers. The resulting one-factor analysis of variance (with identification number of the male as the factor) revealed strong differences among the seven males for which data were available ( $F_{6,2} = 101.8$ ,  $P < 0.01$ ), with matings by two of the males producing 100% inviable offspring, 67% in two others and 29–33% stillborn offspring in litters sired by the other three males. Multiple mating results in shared paternity of the resulting litter<sup>18</sup>, and our data suggest that the ova of females that mate more frequently are fertilized mostly by the sperm of 'better' males. If this were not the case, a higher number of copulations might reduce the variance in proportions of inviable offspring in a litter, but should not affect the mean proportion of inviable neonates. Instead, our data show that the proportion of viable offspring increases with additional matings.

Why should a higher number of matings per litter result in fertilization of a female's ova by a 'better' male? Multiple mating increases the potential for intrauterine competition among sperm from different males. If sperm that are more successful in fertilizing the ova are also more effective in producing viable offspring (as shown in some inbred populations of mammals<sup>21</sup>), then multiple mating should increase the average viability of offspring<sup>11–13</sup>. Such an effect would normally be difficult to detect because of the problems inherent in assessing offspring quality, and the high proportion of stillbirths in the Smygehuk population overcomes this problem. Nonetheless, the same effect may well be widespread among other populations of adders (and many other types of animals). The enhancement of offspring quality through intrauterine sperm competition offers a plausible selective advantage for the willingness of females of many species to mate frequently, with many partners. □

- Manhagen, C. *Trends Ecol. Evol.* **6**, 183–186 (1991).
- Bateman, A. J. *Heredity* **2**, 349–368 (1948).
- Williams, G. C. *Adaptation and Natural Selection: a Critique of Some Current Evolutionary Thought* (Princeton University Press, New Jersey, 1966).
- Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
- Ridley, M. *Biol. Rev.* **63**, 509–549 (1988).
- Turner, M. E. & Anderson, W. W. *Evolution* **37**, 714–723 (1983).
- Parker, G. A. *Biol. Rev.* **45**, 525–567 (1970).
- Davies, E. M. & Boersma, P. D. *Am. Nat.* **123**, 594–611 (1984).
- Stacey, P. B. *Am. Nat.* **120**, 51–64 (1982).
- Devine, M. C. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 509–521 (Academic, Orlando, Florida, 1984).
- Parker, G. A. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 1–60 (Academic, Orlando, Florida, 1984).
- Loman, J., Madsen, T. & Håkansson, T. *Oikos* **52**, 69–72 (1988).
- Shine, R. in *Biology of the Reptilia* Vol. 16 (eds Gans, C. & Huey, R. B.) 275–330 (Liss, New York, 1988).
- Saint Girons, H. & Kramer, E. *Revue suisse Zool.* **70**, 191–221 (1963).
- Nilson, G. *Amphibia-Reptilia* **2**, 63–82 (1981).
- Stille, B., Madsen, T. & Niklasson, M. *Oikos* **47**, 173–175 (1986).
- Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, Cambridge, 1978).
- Madsen, T. & Shine, R. *Evolution* (in the press).
- Ralls, K., Brugger, K. & Ballou, J. *Science* **206**, 1101–1103 (1979).

ACKNOWLEDGEMENTS. We thank E. L. Charnov, M. J. Caley and C. R. Dickman for comments and the Australian Research Council for financial support.

## Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors

Xiao-kun Zhang, Birgit Hoffmann, Paul B.-V. Tran, Gerhart Graupner & Magnus Pfahl\*

Cancer Center, La Jolla Cancer Research Foundation, La Jolla, California 92037, USA

THYROID hormones and retinoic acid function through nuclear receptors that belong to the steroid/thyroid-hormone receptor superfamily (reviewed in refs 1–4). Thyroid hormone receptors (TRs) and retinoic acid receptors (RARs) require auxiliary nuclear proteins for efficient DNA binding<sup>5–10</sup>. Here we report that retinoid X receptors RXR $\alpha$  (ref. 11) is one of these nuclear proteins. RXR $\alpha$  interacts both with TRs and with RARs, forming heterodimers in solution that strongly interact with a variety of T3/retinoic acid response elements. Transfection experiments show that RXR $\alpha$  can greatly enhance the transcriptional activity of TR and RAR at low retinoic acid concentrations that do not significantly activate RXR $\alpha$  itself. Thus, RXR $\alpha$  enhances the transcriptional activity of other receptors and its own ligand sensitivity by heterodimer formation. Our studies reveal a new subclass of receptors and a regulatory pathway controlling nuclear receptor activities by heterodimer formation.

We investigated the possibility that the auxiliary nuclear proteins might be members of the nuclear receptor proteins, in particular those that bind and activate the same or related response elements. Using a gel retardation assay, we observed that TR $\alpha$  DNA binding greatly increases in the presence of RXR $\alpha$ . A prominent complex which migrated much more slowly than the monomeric TR $\alpha$  complex<sup>5</sup> was observed, whereas TR $\alpha$  binding was reduced (Fig. 1a). At the concentrations used, RXR $\alpha$  alone did not form visible complexes with the thyroid hormone response element (TRE). By contrast, there was no change of TR $\alpha$  binding when mixed with RAR $\alpha$  (Fig. 1a) or oestrogen receptor (ER) (Fig. 1b). In addition, when RXR $\alpha$  was mixed with oestrogen receptor and labelled oestrogen response element (ERE) (Fig. 1b), no increased binding or slow electrophoretic mobility complex was seen. When the carboxy-terminal variant TR $\alpha$ -2 (ref. 12), which is not a transcriptional activator<sup>13–15</sup>, was used, a low-electrophoretic-mobility complex did not form either (Fig. 1a).

Received 23 September; accepted 14 November 1991.

- Drummond, B. A. III in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 291–370 (Academic, Orlando, Florida, 1984).
- Hrdy, S. B. in *Feminist Approaches to Science* (ed. Bleier, R.) 119–146 (Pergamon, Oxford, 1986).

\* To whom correspondence should be addressed.



RXR $\alpha$  was also mixed with TR $\beta$ , RAR $\alpha$ , RAR $\beta$  and RAR $\gamma$  receptor proteins (Fig. 1a). As with TR $\alpha$ , the complex formed between TR $\beta$  and TRE was upshifted with respect to migration by RXR $\alpha$ . With RAR $\alpha$ , RAR $\beta$  and RAR $\gamma$ , specific protein-DNA complexes that migrated more slowly than the nonspecific complex formed only in the presence of RXR $\alpha$ , whereas RARs alone did not form detectable complexes with the TRE under the conditions used. The mobility of the slow migrating complexes formed between RXR $\alpha$  and TR $\alpha$  was very similar to that formed between TR $\alpha$  and nuclear protein(s) (Fig. 1c) from CV-1 cells<sup>5</sup>, suggesting that the protein(s) found in CV-1 cells might be RXR $\alpha$  or related protein(s).

Gel retardation assays were done using several TRE-related sequences: an inverted repeat of the TRE (TRE/OP); a TRE half-site; the  $\beta$ RARE, a retinoic acid response element<sup>16,17</sup> and the ERE<sup>18</sup> (Fig. 2a). RXR $\alpha$  alone did not bind to these DNA sequences at the concentration we used (data not shown). But

when it was mixed with TR $\alpha$ , a specific slow migrating complex was observed on the TRE and the TRE/OP (Fig. 2b). Binding of TR $\alpha$  to the half-site and ERE was roughly as efficient as to the palindromic TRE in the absence of RXR $\alpha$ . But this binding was abolished in the presence of RXR $\alpha$ , suggesting that a TR $\alpha$ -RXR $\alpha$  complex is formed in solution which has reduced affinity or does not bind the TRE half-site or the ERE. With the  $\beta$ RARE, TR $\alpha$  only formed a weak complex in the presence of RXR $\alpha$  but RAR $\alpha$  and RXR $\alpha$  formed a very strong complex with the  $\beta$ RARE. Both TR $\alpha$  and RAR $\alpha$  alone did not form detectable complexes with  $\beta$ RARE at the concentration used (data not shown). This difference in specificity of the two different complexes is interesting because TR and RAR alone appear to bind equally well to the  $\beta$ RARE<sup>16</sup>. Together these data demonstrate that although RXR $\alpha$  greatly enhances receptor binding to quite different response elements, a dimeric recognition sequence must be present for effective heterodimer binding

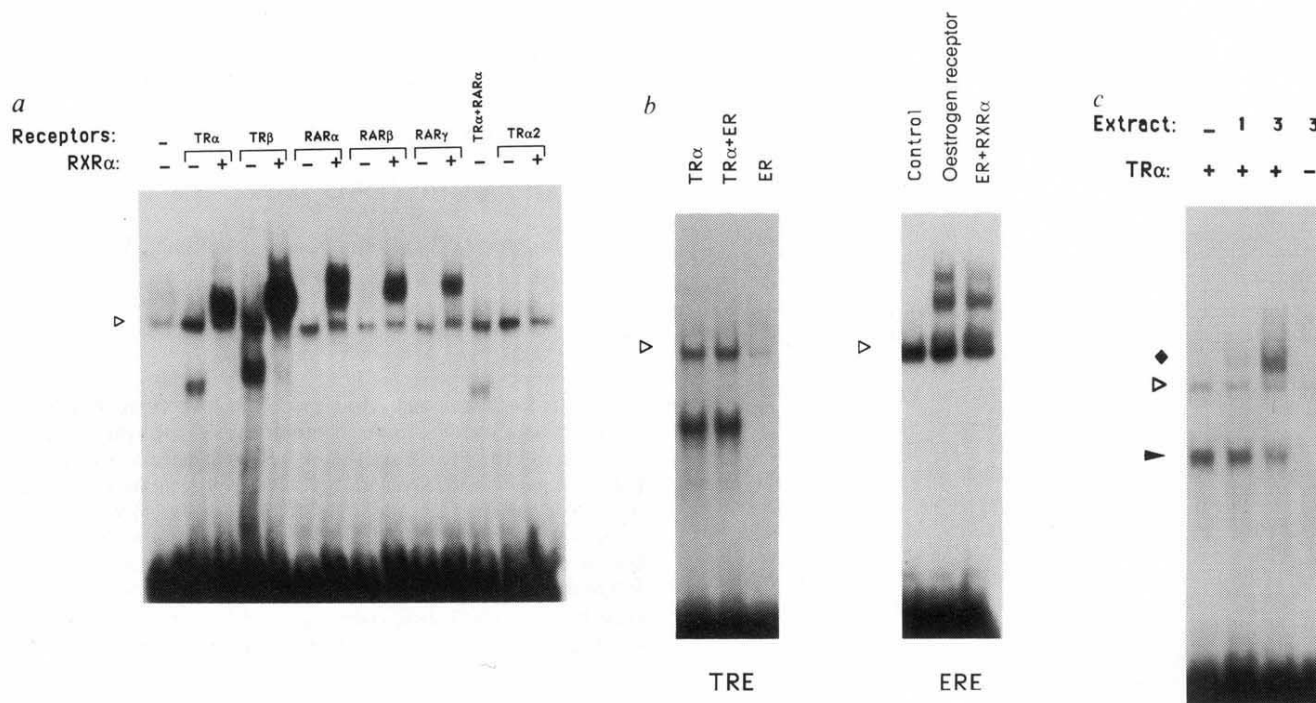


FIG. 1 Enhancement of TR and RAR DNA binding by RXR $\alpha$ . *a*, *In vitro* synthesized TR $\alpha$ , TR $\beta$ , RAR $\alpha$ , RAR $\beta$ , RAR $\gamma$  and TR $\alpha$ 2 receptor proteins were preincubated either with (+) or without (–) equal amount of *in vitro* synthesized RXR $\alpha$  protein at room temperature for 10 min. After this preincubation, the reaction mixtures were incubated with <sup>32</sup>P-labelled palindromic TRE (for sequence see Fig. 2a) and analysed by gel retardation assay as described<sup>5</sup>. Lane 1, binding of unprogrammed reticulocyte lysate. Open arrow, nonspecific band observed with unprogrammed reticulocyte lysate. Complexes migrating below or above the nonspecific band are the specific complexes of TRs in the absence of RXR $\alpha$  and the complexes formed by TRs and RARs in the presence of RXR $\alpha$ . As a control, equal amounts of TR $\alpha$  and RAR $\alpha$  proteins were also mixed and incubated with labelled TRE. *b*, Effect of oestrogen receptor. To analyse whether oestrogen receptor could also enhance TR $\alpha$  binding to the TRE, or whether RXR $\alpha$  would enhance oestrogen receptor binding to the ERE, equal amounts of *in vitro* synthesized oestrogen receptor protein were incubated with TR $\alpha$  or RXR $\alpha$  proteins and the reaction mixtures were analysed by gel retardation using either <sup>32</sup>P-labelled palindromic TRE or palindromic ERE as indicated. Control, binding of the unprogrammed reticulocyte to ERE. Open arrows, nonspecific bands observed with unprogrammed lysate. *c*, Effect of CV-1 cell extract on TR $\alpha$  DNA binding. Cell extract was prepared from CV-1 cells as described<sup>5</sup> and the different amounts of cell extract (in  $\mu$ g) were incubated either with *in vitro* synthesized TR $\alpha$  protein or the same volume of unprogrammed reticulocyte lysate. The reaction mixtures were then analysed by gel retardation using <sup>32</sup>P-labelled palindromic TRE. Open triangle, solid arrow, and solid diamond, nonspecific binding of the reticulocyte lysate, specific TR $\alpha$  binding, and the upshifted

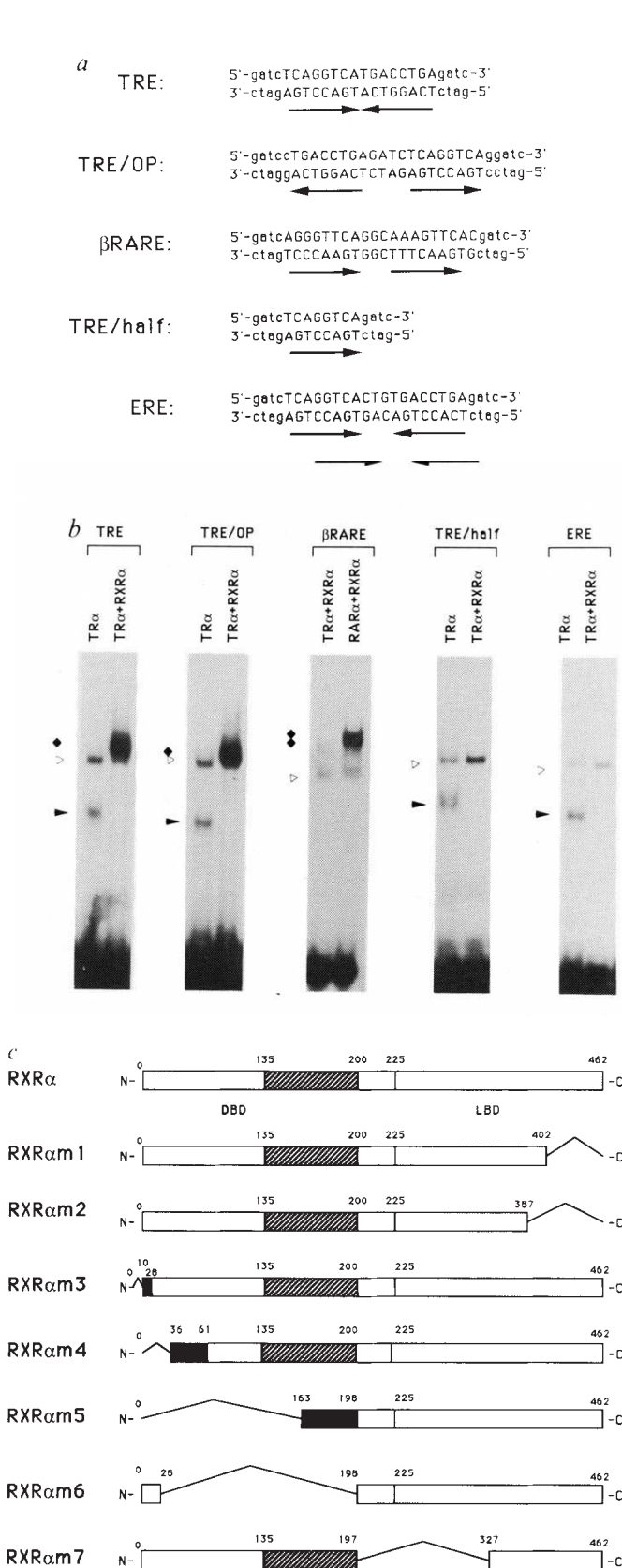
TR $\alpha$  complex, respectively.

**METHODS.** Two primers (5'-CGCAGACATGGACACCAACAT-3'; 5'-CCT-CTCCACCGCATGTCCTCG-3') were used to amplify the N-terminal half of RXR $\alpha$  complementary DNA from SCC-13 by polymerase chain reaction (PCR) technique. The *Sma*1-*Sal*1 fragment from PCR product (530 base pairs (bp)) containing the DNA-binding domain of the RXR $\alpha$  was used as a probe to isolate RXR $\alpha$  cDNA by screening a  $\lambda$ gt11 human placenta cDNA library (obtained from J. Millán; ref. 25). Several positive clones were obtained, including full-length receptor and the truncated clones, RXR $\alpha$ m3, RXR $\alpha$ m4 and RXR $\alpha$ m5 which were sequenced and show identical sequences as the wild-type RXR $\alpha$ . The cDNA clones were then subcloned into the *Eco*RI site of pBluescript and pECE. To obtain receptor proteins, receptor cDNAs were transcribed and translated *in vitro* as described<sup>5,21</sup>. Cell extracts were prepared from CV-1 cells in a buffer containing 20 mM HEPES, pH 7.9, 0.4 M KCl, 2 mM DTT and 20% glycerol as described<sup>5</sup>. For gel retardation assay, *in vitro* translated receptor protein (1–5  $\mu$ l depending on the translation efficiency) was incubated with the <sup>32</sup>P-labelled oligonucleotides in a 20- $\mu$ l reaction mixture containing 10 mM HEPES buffer, pH 7.9, 50 mM KCl, 1 mM DTT, 2.5 mM MgCl<sub>2</sub>, 10% glycerol and 1  $\mu$ g of poly(dI-dC) at 25 °C for 20 min. The reaction mixture was then loaded on a 5% nondenaturing polyacrylamide gel containing 0.5  $\times$  TBE (1  $\times$  TBE is 0.089 M Tris-borate, 0.089 M boric acid and 0.002 M EDTA). To analyse the effect of RXR $\alpha$  or the nuclear proteins on receptor DNA binding activity, RXR $\alpha$  or the cell extracts were preincubated with receptor protein at room temperature for 10 min before the DNA binding assay.

and in addition that RXR greatly enhances the specificity of TRs or RARs for particular responsive elements.

Two regions of TR $\alpha$  seem to be needed for dimerization and interaction with nuclear auxiliary proteins<sup>5</sup>. One region includes part of the DNA-binding domain and the other region includes the heptad repeat region<sup>19</sup> in the carboxy-terminal domain<sup>5</sup>. To

delineate regions of RXR $\alpha$  involved in nuclear receptor interaction, we examined deletion mutants of RXR $\alpha$  (Fig. 2c,d). Deletion of 60 (RXR $\alpha$ m1) or 75 (RXR $\alpha$ m2) amino acids from the RXR $\alpha$  C terminus abolished enhancement and upshift of the TR $\alpha$  band whereas deletion of 28 (RXR $\alpha$ m3) or 61 (RXR $\alpha$ m4) amino acids from the amino terminus did not visibly



**FIG. 2** Interaction of TR $\alpha$  and RXR $\alpha$ . **a**, Sequences of oligonucleotides used for the gel retardation assays. TRE is the perfect palindromic T3/retinoic acid response element<sup>20,21</sup>. TRE/OP, an oligonucleotide consisting of two TRE half-site (as indicated by the arrows) in the opposite orientation separated by 4 bp<sup>29</sup>. βRARE, a retinoic acid response element present in the RARβ promoter<sup>16</sup>. TRE/half, half-site of TRE. ERE, perfect palindromic oestrogen receptor response element<sup>18</sup>. These oligonucleotides were synthesized with appropriate restriction sites at both ends as indicated by the small letters. **b**, Gel retardation analysis of RXR $\alpha$ -TR $\alpha$  or RXR $\alpha$ -RAR $\alpha$  interaction using different DNA response elements. Gel retardation assays were done using equal amounts of receptor protein essentially as described in Fig. 1a but using different response elements. Specific binding of TR $\alpha$  to each response element is indicated by the solid arrows. Nonspecific bands observed with unprogrammed reticulocyte lysate are indicated by open arrows. Heterodimer formation is indicated by the diamonds. **c**, Schematic representation of the RXR $\alpha$  deletion mutants. RXR $\alpha$  deletion mutants were constructed and proteins were prepared as described below. Numbers above the bars indicate the amino-acid positions. DNA-binding domain (DBD) and ligand-binding domain (LBD) are indicated. Single lines mark the deleted portions of the receptor. Deletion mutants RXR $\alpha$ m3, RXR $\alpha$ m4 and RXR $\alpha$ m5 are truncated cDNA clones of RXR $\alpha$  isolated from screening the human placenta  $\lambda$ gt11 cDNA library (see legend to Fig. 1). These clones were sequenced and show identical nucleotide sequence as the wild-type RXR $\alpha$ . The proteins of these cDNA clones were translated *in vitro* using existing Met codons for amino acid 28, 61 and 198, respectively, as determined by SDS-PAGE. The black bars indicate the untranslated portions of these three mutants. **d**, Interaction of RXR $\alpha$  deletion mutants with TR $\alpha$ . Interaction of RXR $\alpha$  deletion mutants with TR $\alpha$  was analysed by the gel retardation assay essentially as described in Fig. 1a. The first lane represents the binding of the unprogrammed reticulocyte lysate. The nonspecific binding is indicated by the open arrow. The specific band of TR $\alpha$  migrates faster than the nonspecific band. Arrows indicate the migration positions of complexes formed with RXR $\alpha$ m3 and RXR $\alpha$ m4.

**METHODS.** RXR $\alpha$  deletion mutants were obtained as follows: RXR $\alpha$ m1 and RXR $\alpha$ m2 were generated by digesting RXR $\alpha$  cDNA with *Stu*I (1,463) and *Xma*III (1,231), respectively. RXR $\alpha$ m6 and RXR $\alpha$ m7 were generated by internal deletion using *Nco*I and *Bal*I, respectively. RXR $\alpha$ m3, RXR $\alpha$ m4 and RXR $\alpha$ m5 are truncated RXR $\alpha$  cDNA clones. When sequenced they showed sequences identical to wild-type RXR $\alpha$ .



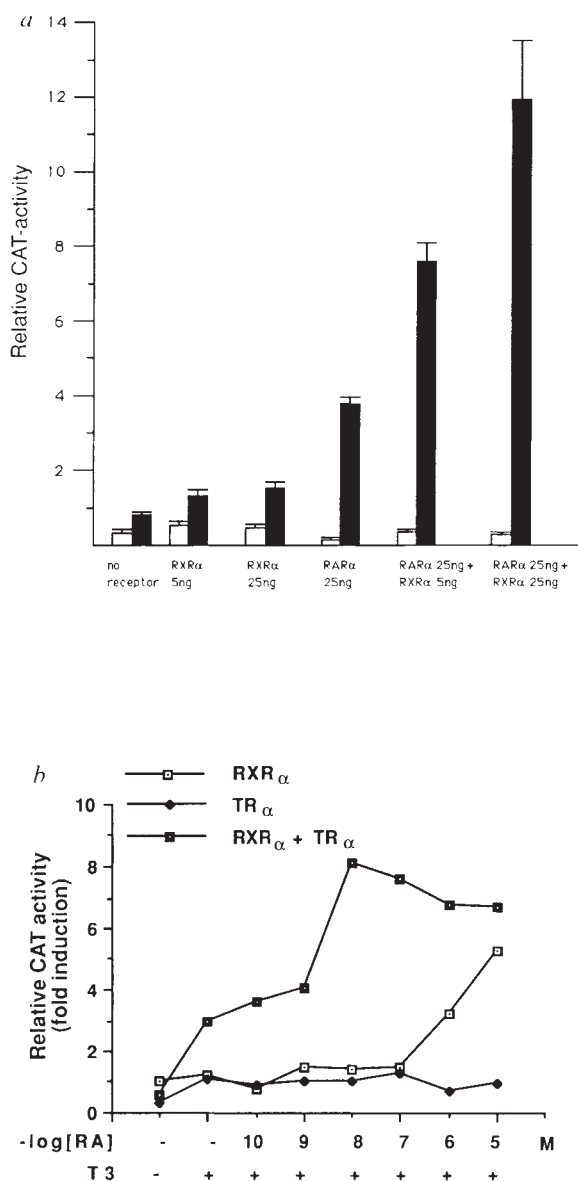


FIG. 3 RXR $\alpha$  enhances the transcriptional activation of RARs and TR. *a*, CV-1 cells were cotransfected with 100 ng TRE<sub>2</sub>-tk-CAT and the indicated amounts of the receptor expression vector. Cells were treated with 100 nM retinoic acid (■) or no hormone (□), and 24 h later assayed for CAT activity. The mean of duplicate cultures is shown. *b*, Induction curves of RXR $\alpha$  in the presence and absence of TR $\alpha$ . The palindromic TRE reporter gene (TRE-tk-CAT, 100 ng per well) was transfected into CV-1 cells together with 5 ng RXR $\alpha$ , 150 ng  $\beta$ -galactosidase plasmid, 25 ng TR $\alpha$  (or no TR $\alpha$ ) and Bluescript up to 1,000 ng. Cells were grown in 24-well plates with the indicated concentrations of retinoic acid ([RA]) and a constant amount of  $10^{-7}$  M T<sub>3</sub>. CAT activities were corrected for transfection efficiency by  $\beta$ -galactosidase values. As control, reporter constructs were transfected alone, and CAT activities were analysed after the same hormone treatment as described above. The activity of RXR $\alpha$  on the reporter gene in the absence of hormone was chosen as the reference value, and CAT activities were normalized accordingly. The mean values from four to six independent transfection experiments are shown.

**METHODS.** CV-1 cells were grown in DME medium supplemented with 10% fetal calf serum (FCS). Cells were plated at  $1.0 \times 10^5$  per well in a 24-well plate 16–24 h before transfection as described previously<sup>26</sup>. A modified calcium phosphate precipitation procedure was used for transient transfection and is described elsewhere<sup>27</sup>. In general, 100 ng of reporter plasmid, 150 ng  $\beta$ -galactosidase expression vector (pCH 110, Pharmacia), and variable amounts of receptor expression vector were mixed with carrier DNA (Bluescript) to 1,000 ng of total DNA per well. CAT activity was normalized for transfection efficiency by the corresponding  $\beta$ -galactosidase activity<sup>27</sup>. The receptor expression vectors have been described<sup>21,27–29</sup>.

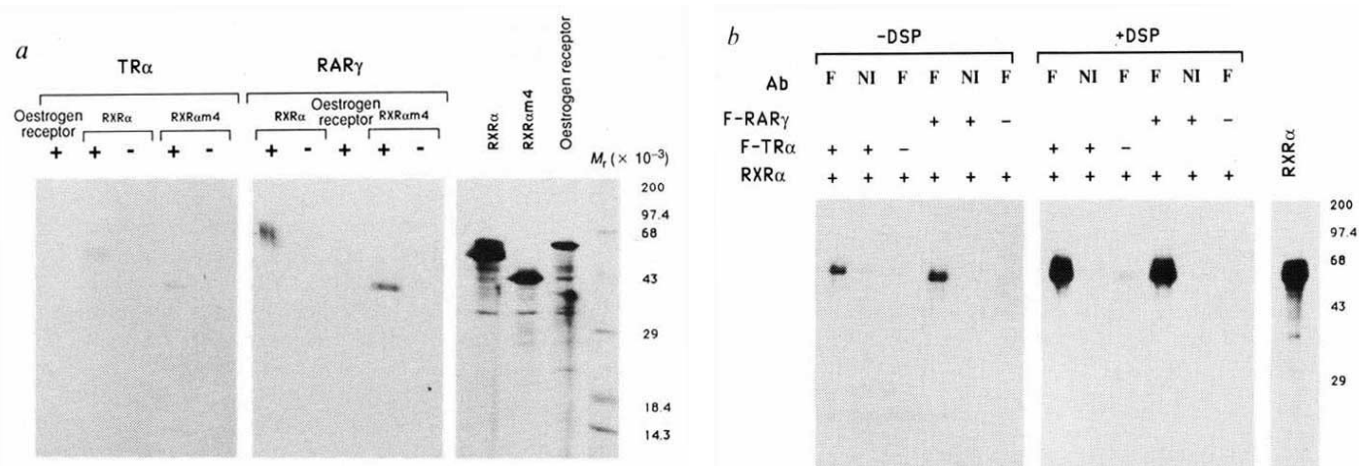
affect interaction with TR $\alpha$  (Fig. 2*d*). But a comparison of the complexes with RXR $\alpha$ m3 and RXR $\alpha$ m4 (indicated by arrows) shows that the size of the RXR protein determines migration of the complex, providing evidence that RXR $\alpha$  participates in the complex. We also observed that the C-terminal truncation of TRs and RARs impaired their interaction with RXR $\alpha$  (data not shown). An internal deletion spanning the hinge region and the N-terminal half of the ligand-binding domain (RXR $\alpha$ m7) also abolished interaction with TR $\alpha$  (Fig. 2*d*). Together, these results suggest that for TRs as well as for RXR $\alpha$  the C-terminal domains have a role in heterodimer formation. But two RXR $\alpha$  mutants, RXR $\alpha$ m5 and RXR $\alpha$ m6, which both lacked portions of the DNA-binding domain, also failed to form a complex with TR $\alpha$  on the TRE. The absence of a complex and that TR $\alpha$  DNA binding was not inhibited may suggest that portions of the DNA-binding region of RXR $\alpha$  are required for interaction with TR $\alpha$ . When we replaced TR $\alpha$  with TR $\beta$ , RAR $\alpha$ , RAR $\beta$  or RAR $\gamma$ , identical results were obtained with the RXR $\alpha$  mutants (data not shown).

The ability of RXR $\alpha$  to enhance RAR and TR DNA binding could also allow enhancement of transcriptional activation of these receptors on the TRE, a known T<sub>3</sub> and retinoic acid response element<sup>20,21</sup>. When low concentrations of RXR $\alpha$  expression vector were cotransfected with RAR $\alpha$  and the TRE<sub>2</sub>-tk-CAT reporter construct, a strong enhancement of the RAR $\alpha$  activity was observed (Fig. 3*a*). This strong enhancing activity was at retinoic acid concentrations ( $10^{-7}$  M) at which RXR $\alpha$  alone was only slightly activated<sup>11</sup>. The increased activation of the reporter gene in the presence of both retinoid receptors can be more than additive at certain receptor concentrations as shown (Fig. 3*a*).

To determine whether RXR $\alpha$  itself requires ligand binding to boost transcriptional activation, we cotransfected RXR $\alpha$  with TR $\alpha$ . A strong synergism was already seen when only thyroid hormone (T<sub>3</sub>) was added, whereas an optimal reporter gene response required the presence of both ligands, T<sub>3</sub> and retinoic acid (Fig. 3*b*). Only small amounts of RXR $\alpha$  as well as TR $\alpha$  were required for a strong activation of the reporter genes. There was a large shift of the retinoic acid responsiveness of RXR $\alpha$  in the presence of TR $\alpha$  in that the RXR $\alpha$  sensitivity to retinoic acid seemed to be increased by at least 100 times (Fig. 3*b*). This enhanced ligand sensitivity is not due to the activation of endogenous RARs by TR $\alpha$  because no effect of CAT activity was observed when TR $\alpha$  was transfected alone (Fig. 3*b*) consistent with previous observations<sup>21</sup>. Thus, the ligand responsiveness of RXR $\alpha$  also seems to be greatly increased in the heterodimer complex.

An important question is whether RXR $\alpha$  can form heterodimers with the other receptors in solution or whether the heterodimeric complexes are only formed in the presence of specific DNA sequences. TR $\alpha$ - or RAR $\gamma$ -glutathione transferase fusion proteins were produced in bacteria, bound to a glutathione-containing affinity resin<sup>22</sup> and were then mixed with *in vitro* synthesized <sup>35</sup>S-labelled receptors. Only labelled RXR $\alpha$  could be specifically eluted with glutathione from columns that contained bound TR $\alpha$  or RAR $\gamma$  fusion protein, but not labelled oestrogen receptor, whereas the mutant RXR $\alpha$ m4 that lacked 61 amino acids at the N terminus and was able to upshift the migration of TR $\alpha$  and RAR $\gamma$  was also retained (Fig. 4*a*).

We then incubated labelled RXR $\alpha$  protein with or without TR $\alpha$  or RAR $\gamma$  proteins that contained an eight amino-acid epitope (Flag) at the N-terminal end of these receptors (Flag-TR $\alpha$  and Flag-RAR $\gamma$ ). Anti-Flag antibody was used to examine whether RXR $\alpha$  could be precipitated together with Flag-TR $\alpha$  or Flag-RAR $\gamma$ . Precipitation of Flag-TR $\alpha$  or Flag-RAR $\gamma$  resulted in a significant specific coprecipitation of labelled RXR $\alpha$  protein, specific coprecipitation was largely enhanced when cross-linker (DSP) was used (Fig. 4*b*). These immuno coprecipitation data together with results obtained with the affinity



**FIG. 4** Direct interaction of RXR $\alpha$  with TR or with RAR. **a**, Affinity column chromatography. To analyse whether RXR $\alpha$  directly interacts with TRs and with RARs in the absence of DNA, TR $\alpha$  and RAR $\gamma$  proteins were synthesized in bacteria using PGEX-2T expression vector (Pharmacia). Purified glutathione S-transferase-TR $\alpha$  or -RAR $\gamma$  fusion proteins were bound to the glutathione sepharose column. As a control, the same amount of glutathione transferase protein was also bound to a column (-). <sup>35</sup>S-labelled RXR $\alpha$  and the mutant RXR $\alpha$ 4 synthesized *in vitro* were then loaded on columns that contained bound glutathione transferase-TR $\alpha$  or -RAR $\gamma$  or glutathione transferase. As a control, *in vitro* synthesized <sup>35</sup>S-labelled oestrogen receptor was also loaded on a column containing bound glutathione transferase-TR $\alpha$  or -RAR $\gamma$ . After extensive washing with PBS, the bound proteins were eluted with 5 mM reduced glutathione. The eluates were concentrated using centricon 10 and analysed on a 10% SDS-PAGE. The right panel represents *in vitro* translation products of RXR $\alpha$ , RXR $\alpha$ 4, and ER. **b**, Immunoprecipitation of RXR $\alpha$  by antibody against TR or RAR. <sup>35</sup>S-labelled *in vitro* synthesized RXR $\alpha$  protein was incubated with partially purified bacterially expressed Flag-TR $\alpha$  or Flag-RAR $\gamma$  (+) or similarly prepared glutathione transferase control protein (-) either in the absence or presence of cross-linker DSP as indicated. After incubation, either anti-Flag antibody (F) or preimmune serum (NI) was added. The immune complexes were washed, boiled in SDS sample buffer and separated on a 10% SDS-PAGE. The labelled, *in vitro* synthesized RXR $\alpha$  protein is shown in the right panel together with molecular mass markers (M<sub>r</sub> × 10<sup>-3</sup>).

**METHODS.** The construction of Flag-containing receptors (Flag-TR $\alpha$  and Flag-RAR $\gamma$ ) has been described previously<sup>5,15</sup>. Briefly, they were constructed by ligating a double-stranded oligonucleotide containing an ATG codon and a DNA sequence encoding Flag (Arg Tyr Lys Asp Asp Asp Asp Lys) to the N-terminus of receptors. To prepare TR $\alpha$  and RAR $\gamma$  fusion proteins, Flag-TR $\alpha$  and Flag-RAR $\gamma$  cDNAs were cloned in frame into the expression vector pGex-2T (Pharmacia). The proteins were expressed in bacteria using the

procedure provided by the manufacturer. Proteins were purified on a pre-packed glutathione Sepharose 4B column (Pharmacia), and checked by gel retardation assays and western blot with anti-Flag antibody. To analyse the interaction between RXR $\alpha$  and TR $\alpha$  or and RAR $\gamma$ , purified Flag-TR $\alpha$  or Flag-RAR $\gamma$  fusion proteins were loaded on the prepacked glutathione Sepharose 4B columns. For control, the vector protein (glutathione S-transferase) prepared under the same conditions was also loaded on the separate columns. The columns were washed extensively with PBS with 1% Triton X-100. <sup>35</sup>S-labelled *in vitro* synthesized RXR $\alpha$ , RXR $\alpha$ 4 and ER proteins were applied to the columns. Columns were then washed extensively three times with 10 ml PBS. The bound protein was eluted with 50 mM Tris, pH 8, containing 5 mM glutathione. Eluates were then concentrated by using a Centricon 10 microconcentrator, and analysed by denaturing polyacrylamide gels. For immunoprecipitation assay, 20  $\mu$ l reticulocyte lysate containing *in vitro* translated <sup>35</sup>S-labelled RXR $\alpha$  were incubated with 5  $\mu$ l (roughly 0.2  $\mu$ g) of partially purified bacterially expressed Flag-TR $\alpha$  or Flag-RAR $\gamma$  fusion proteins or similarly prepared glutathione transferase control protein in 100  $\mu$ l buffer (containing 50 mM KCl and 10% glycerol) for 15 min at room temperature. When cross-linker was used, we added 2  $\mu$ l 100 mM Dithiobis(succinimidylpropionate) (DSP) and continued the incubation at room temperature for 10 min. The reactions were then incubated with 1  $\mu$ l anti-Flag antibody or preimmune serum for 2 h on ice. Immunocomplexes were precipitated by adding 60  $\mu$ l protein-A-Sepharose slurry and mixing continuously in the cold room for 1 h. Protein-A-Sepharose was saturated in TBS buffer (Tris-buffered saline) or in radioimmunoprecipitation assay (RIPA) buffer when crosslinker was used. The immunocomplexes were washed four times with NET-N buffer (20 mM Tris, pH 8.0, 100 mM NaCl, 1 mM DTT, 0.5% Nonidet P-40) or five times with RIPA buffer when DSP was used, and resuspended in SDS sample buffer containing 15%  $\beta$ -mercaptoethanol, boiled and resolved by SDS-PAGE. The gels were fixed, dried and visualized by autoradiography.

column strongly suggest that RXR $\alpha$  forms a stable complex with TRs and RARs in solution.

We provide here evidence that RXR $\alpha$  enhances DNA binding and DNA specificity of TRs or RARs to a number of T<sub>3</sub>/retinoic acid dimeric response elements. Interaction of RXR $\alpha$  with the other receptors occur in solution and seems to require the C-terminal region as well as part of the DNA-binding domain. The enhancement of DNA binding in the presence of RXR $\alpha$  is very similar to the enhanced DNA binding of TR $\alpha$  observed in the presence of nuclear extract from several cell lines<sup>5</sup>. Some TR $\alpha$  mutants also behaved almost identically to RXR $\alpha$  and nuclear extract protein<sup>5</sup> (data not shown). Thus RXR $\alpha$  is probably identical or closely related to cellular protein activities previously described<sup>5,7-10</sup>. According to their enhancing effects on TR DNA binding, those nuclear protein(s) have been termed TR auxiliary proteins, TRAP (ref. 6). But this is insufficient to describe RXR function. Our data suggest that one of the major roles of RXR $\alpha$  is to boost the activity of TRs and RARs in terms of DNA binding specificity and transcriptional activation, while at the same time also enhancing its own effectiveness by

heterodimerization, which we term 'booster receptor' (B-receptor), in contrast to activator receptors (A-receptors).

That the RXR/TR $\alpha$  complex seems to require two distinct hormones for full activation suggests direct cross-talk between two different hormonal signals at the receptor level. An example for this may be the growth hormone gene where T<sub>3</sub> increases the effectiveness of retinoic acid by about 1,000 times<sup>23</sup>. Our finding that in the heterodimer only low retinoic acid concentrations (10<sup>-8</sup>-10<sup>-7</sup> M) are required for full activation of RXR $\alpha$ , may indicate that retinoic acid is an important natural ligand for RXR $\alpha$ . Apart from different RXRs<sup>11,24</sup>, the subfamily of B-receptors may also include a substantial number of orphan receptors for which no high-affinity ligands could be detected. □

Received 29 November; accepted 31 December 1991.

- Green, S. & Chambon, P. *Trends Genet.* **4**, 309-314 (1988).
- Evans, R. M. *Science* **240**, 889-895 (1988).
- Pfahl, M. et al. in *Retinoids in Normal Development and Teratogenesis* (ed. G. Morriss-Kay) 51-64 (Oxford University Press, 1992).



4. Glass, C. K. & Holloway, J. M. *Biochim. biophys. Acta* **1032**, 157–176 (1990).
5. Zhang, X.-K., Tran, P. B.-V. & Pfahl, M. *Molec. Endocr.* **5**, 1909–1920 (1991).
6. Rosen, E. D., O'Donnell, A. L. & Koenig, R. J. *Molec. cell. Endocr.* **78**, C83–C88 (1991).
7. Murray, M. B. & Towle, H. C. *Molec. Endocr.* **3**, 1434–1442 (1989).
8. Lazar, M. A. & Berrodrin, T. J. *Molec. Endocr.* **4**, 1627–1635 (1990).
9. Glass, C. K., Devary, O. V. & Rosenfeld, M. G. *Cell* **63**, 729–738 (1990).
10. Burnside, J., Darling, D. S. & Chin, W. W. *J. biol. Chem.* **265**, 2500–2504 (1990).
11. Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. *Nature* **345**, 224–229 (1990).
12. Benbrook, D. & Pfahl, M. *Science* **238**, 788–791 (1987).
13. Koenig, R. J. *et al. Nature* **337**, 659–660 (1989).
14. Schueler, P. A., Schwartz, H. L., Strait, K. A., Mariash, C. N. & Oppenheimer, J. H. *Molec. Endocr.* **4**, 227–234 (1990).
15. Hermann, T. *et al. Cell Regulation* **2**, 565–574 (1991).
16. Hoffmann, B. *et al. Molec. Endocr.* **4**, 1727–1736 (1990).
17. de Thé, H., Vivanco-Ruiz, M. M., Tiollais, P., Stunnenberg, M. & Dejean, A. *Nature* **343**, 177–180 (1990).
18. Klein-Hitpass, L., Schorpp, M., Wagner, U. & Ryffel, G. U. *Cell* **46**, 1053–1061 (1986).
19. Forman, B. M. & Samuels, H. H. *Molec. Endocr.* **4**, 1293–1301 (1990).
20. Glass, C. K., Holloway, J. M., Devary, O. V. & Rosenfeld, M. G. *Cell* **54**, 313–323 (1988).
21. Graupner, G., Wills, K. N., Tzukerman, M., Zhang, X.-K. & Pfahl, M. *Nature* **340**, 653–656 (1989).
22. Smith, D. B. & Johnson, K. S. *Gene* **67**, 31–40 (1988).
23. Bedo, G., Santisteban, P. & Aranda, A. *Nature* **339**, 231–233 (1989).
24. Hamada, K. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8289–8293 (1989).
25. Millán, J. L. *J. biol. Chem.* **261**, 3112–3115 (1986).
26. Husmann, M. *et al. Molec. cell. Biol.* **11**, 4097–4103 (1991).
27. Pfahl, M. *et al. Meth. Enzym.* **18**, 256–270 (1990).
28. Lehmann, J. M., Hoffmann, B. & Pfahl, M. *Nucleic Acids Res.* **19**, 573–578 (1991).
29. Zhang, X.-K. *et al. New Biol.* **3**, 1–14 (1991).

**ACKNOWLEDGEMENTS.** We thank S. Smith-Ortega for preparation of the manuscript, S. Delgado for artwork and M. Pfahl for technical assistance; R. Maki and C. Hauser for reading the manuscript; T. Hermann and M. Tzukerman for discussion and for unpublished material. X.-K.Z. was supported by a postdoctoral fellowship from the Tobacco Related Disease Research Program of the University of California. B.H. was supported by a fellowship from the Deutsche Forschungsgemeinschaft. This work was supported by the NIH and a grant from the Tobacco Related Disease Program of the University of California.

## Retinoid X receptor interacts with nuclear receptors in retinoic acid, thyroid hormone and vitamin D<sub>3</sub> signalling

Steven A. Kliewer\*, Kazuhiko Umesono\*†, David J. Mangelsdorf\* & Ronald M. Evans\*†

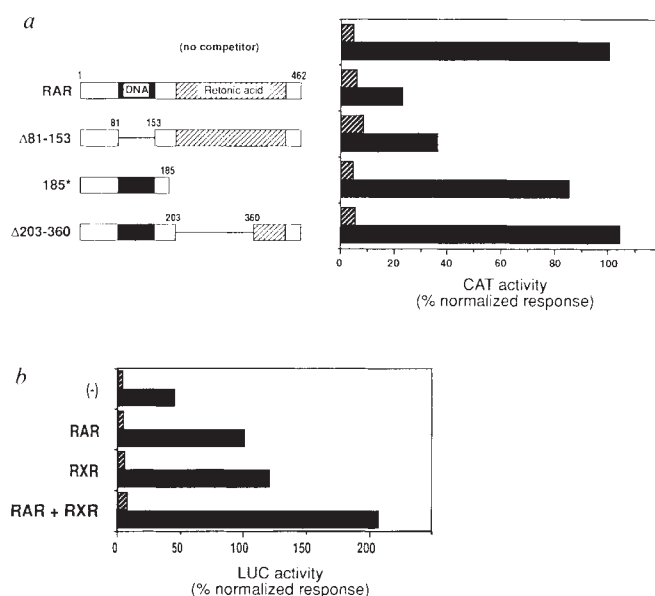
† Howard Hughes Medical Institute and \* The Salk Institute for Biological Studies, La Jolla, California 92037, USA

**CELLULAR responsiveness to retinoic acid and its metabolites is conferred through two structurally and pharmacologically distinct<sup>1</sup> families of receptors: the retinoic acid receptors (RAR)<sup>2,3</sup> and the retinoid X receptors (RXR)<sup>1</sup>. Here we report that the transcriptional activity of RAR and RXR can be reciprocally modulated by direct interactions between the two proteins. RAR and RXR have a high degree of cooperativity in binding to target DNA, consistent with previous reports indicating that the binding of either RAR or RXR to their cognate response elements is enhanced by factors present in nuclear extracts<sup>4,5</sup>. RXR also interacts directly with and enhances the binding of nuclear receptors conferring responsiveness to vitamin D<sub>3</sub> and thyroid hormone T<sub>3</sub>; the DNA-binding activities of these receptors are also stimulated by the presence of nuclear extracts<sup>6–9</sup>. Together these data indicate that RXR has a central role in multiple hormonal signalling pathways.**

RXR, but not RAR, can activate gene expression through the RXR response element found in the promoter of the cellular retinol binding protein type II (CRBP-II-RXRE)<sup>10</sup>. Unexpectedly, RAR represses RXR-mediated activation through the CRBP-II-RXRE (ref. 10 and Fig. 1a). By contrast, transfection of expression plasmids for either RAR or RXR resulted in an induction of expression from a reporter construct driven by two copies of an RAR response element (RARE)<sup>11</sup> (Fig. 1b, RAR or RXR); cotransfection of expression plasmids for both receptors together yielded an enhanced level of expression relative to transfection of either receptor expression plasmid alone (Fig. 1b, RAR + RXR). Together, these data provide evidence for a functional interaction between the two retinoid responsive receptors.

The ability of RAR to interfere with RXR-dependent transactivation through the CRBP-II-RXRE was exploited as an assay to localize potential domains of interaction between RAR and RXR. Deletion of the DNA-binding domain of RAR ( $\Delta$ 81–153) had only a slight effect on the ability of RAR to block RXR function (Fig. 1a). RAR mutants either lacking the C terminus (185\*) or containing a nested deletion in this region ( $\Delta$ 203–360) failed to block RXR-mediated transactivation through the CRBP-II-RXRE (Fig. 1a). Thus, efficient repression of RXR activity requires the C terminus of RAR.

The C termini of RAR and other members of the receptor superfamily contain regions necessary for homo- and heterodimerization<sup>12–14</sup>. To investigate the possibility of a physical interaction between RAR and RXR, immunoprecipitation experiments were done using bacterially expressed RXR and <sup>35</sup>S-methionine-labelled RAR synthesized *in vitro*. Preincubation of RXR and RAR followed by precipitation with RXR-specific antiserum resulted in the efficient coprecipitation of radio-labelled RAR (Fig. 2a, lane 2). By contrast, no RAR was detected when RXR was omitted from the reaction (Fig. 2a, lane 1).

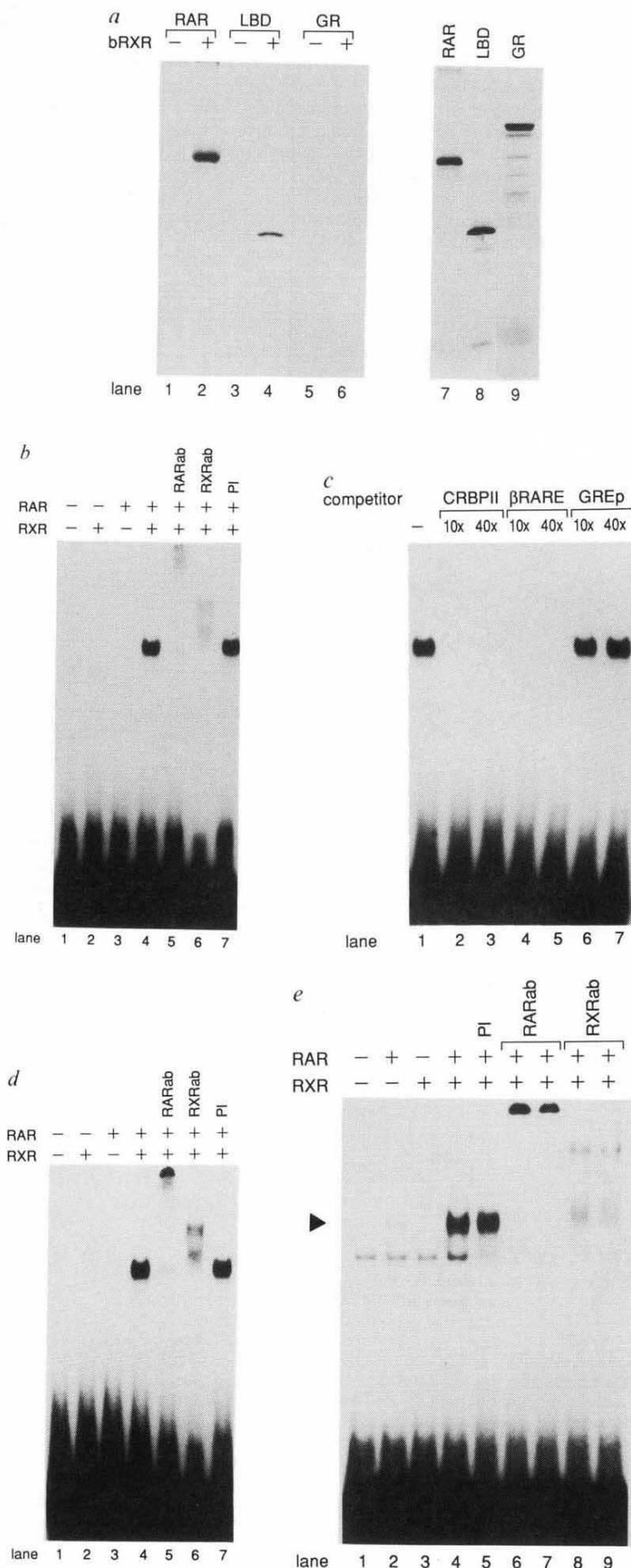


**FIG. 1** *a*, the C terminus of RAR is required for suppression of RXR-dependent transactivation through the CRBP-II-RXRE. CV-1 cells were cotransfected in duplicate with the reporter construct  $\Delta$ SV-CRBP-II-CAT, expression plasmid RS-hRXR $\alpha$ , and the control expression vector RS-LUC (no competitor) or the expression plasmids RS-hRAR $\alpha$ , RS- $\Delta$ 81-153, RS-185\* and RS- $\Delta$ 203-360. Cells (right-hand figure) were then exposed to either ethanol (cross-hatched) or 10  $\mu$ M retinoic acid (filled bars). CAT (chloramphenicol acetyltransferase) activity is presented as per cent conversion where retinoic acid-induced activation in the presence of RXR is arbitrarily set at 100%. *b*, RXR enhances RAR-dependent transactivation through an RARE. CV-1 cells were cotransfected in duplicate with the reporter construct tk-DR5.2-LUC, containing two copies of the DR-5 RARE upstream of the thymidine kinase promoter, and expression plasmids RS-CAT (-), RS-RAR $\alpha$ , and/or RS-RXR $\alpha$  as indicated. Cells were then exposed to either ethanol or 10  $\mu$ M retinoic acid (cross-hatched and filled bars respectively). Luciferase activity is presented as per cent normalized response where retinoic acid-induced activation in the presence of RAR is arbitrarily set at 100%.

**METHODS.** CV-1 monkey kidney cell culture, transfections, CAT and luciferase assays were done as previously described<sup>10,11,20</sup>. Transfection was on 10-cm plates and, for experiments using CAT, included 1  $\mu$ g of RS-receptor or RS-LUC, 0.5  $\mu$ g RS-hRXR $\alpha$ , 1  $\mu$ g  $\Delta$ SV-CRBP-II-CAT reporter, 5  $\mu$ g RAS- $\beta$ -galactosidase reporter and 7.5  $\mu$ g pGEM4 carrier DNA. Transfections using luciferase included 50 ng RS-RAR $\alpha$  and/or 100 ng RXR $\alpha$  (the total amount of RS-expression plasmid was maintained constant in each transfection through the addition of RS-CAT), 0.5  $\mu$ g tk-DR5.2-LUC reporter, 5  $\mu$ g RAS- $\beta$ -galactosidase reporter and 8.5  $\mu$ g pGEM4 carrier DNA.

FIG. 2 Direct interactions between RAR and RXR in the absence or presence of DNA. *a*, RAR and RXR form a complex in solution. Immunoprecipitation reactions were done using *in vitro* synthesized,  $^{35}\text{S}$ -methionine-labelled RAR $\alpha$  (lanes 1 and 2), the ligand binding domain of RAR $\alpha$  (LBD) (amino acids 155–462) (lanes 3 and 4), or the GR (glucocorticoid receptor) (lanes 5 and 6) in either the absence (lanes 1, 3, 5) or presence (lanes 2, 4, 6) of bacterially expressed RXR (bRXR). Polyclonal antisera prepared against RXR $\alpha$  was used. *In vitro* synthesized RAR, LBD, and GR proteins not subjected to immunoprecipitation are shown in lanes 7–9. *b*, RAR and RXR bind cooperatively to the CRBP/II–RXRE. Gel mobility shift assays were done using *in vitro* synthesized RAR and/or RXR as indicated and  $^{32}\text{P}$ -labelled CRBP/II–RXRE oligonucleotide<sup>10</sup>. Polyclonal antisera prepared against either RAR (RARab) (lane 5) or RXR (RXRab) (lane 6) or preimmune serum (PI) (lane 7) were included in the reactions as indicated. *c*, Binding specificity of the RAR–RXR complex. Gel mobility shift competition reactions were done using  $^{32}\text{P}$ -labelled CRBP/II–RXRE oligonucleotide and either a 10-fold (10 $\times$ ) or 40-fold (40 $\times$ ) excess of unlabelled competitor oligonucleotide encoding either the CRBP/II–RXRE (lanes 2 and 3), the  $\beta$ RARE (lanes 4 and 5), or a palindromic GRE<sup>20</sup> (lanes 6 and 7). *d*, RAR and RXR bind cooperatively to the  $\beta$ RARE. Gel mobility shift assays were done using *in vitro* synthesized RAR and/or RXR as indicated and  $^{32}\text{P}$ -labelled  $\beta$ RARE oligonucleotide. Polyclonal antisera prepared against either RAR (RARab) (lane 5) or RXR (RXRab) (lane 6) or preimmune serum (PI) (lane 7) were included in the reactions as indicated. *e*, RAR and RXR overexpressed in COS cells bind cooperatively to the  $\beta$ RARE. Gel mobility shift assays were done using  $^{32}\text{P}$ -labelled  $\beta$ RARE and whole cell extracts prepared from COS cells in which RAR (lane 2), RXR (lane 3) or RAR and RXR (lanes 4–9) were overexpressed. Preimmune serum (3  $\mu\text{l}$ ) (lane 5) or 0.2  $\mu\text{l}$  or 1  $\mu\text{l}$  RAR-specific antiserum (RARab) (lanes 6 and 7, respectively) or 1  $\mu\text{l}$  or 3  $\mu\text{l}$  RXR-specific antiserum (RXRab) (lanes 8 and 9, respectively) were included in the reactions as indicated. The position of the RAR–RXR– $\beta$ RARE complex is indicated by an arrowhead.

**METHODS.** The LBD expression vector was generated through insertion of an *XhoI*–*BamHI* fragment of  $\Delta 81$ –153, including amino acids 155–462 of RAR $\alpha$ , into the pCMX expression vector containing a synthetic translation start site sequence<sup>11,21</sup>. RAR $\alpha$ , LBD, and GR RNA was prepared and subsequently translated in rabbit reticulocyte lysates as directed by the supplier (Promega). RXR was expressed in bacteria as a fusion with glutathione-S-transferase using the pGEX-2T expression vector (Pharmacia) and purified as previously described<sup>10</sup>. Immunoprecipitation reactions (20  $\mu\text{l}$ ) included 5  $\mu\text{l}$   $^{35}\text{S}$ -methionine-labelled receptor protein and 150 ng of either purified GST–RXR or GST alone in 20 mM Tris, pH 8.0. Proteins were incubated 20 min on ice before the addition of 5  $\mu\text{l}$  polyclonal antisera prepared against an RXR $\alpha$  peptide (amino acids 214–229). Antigen–antibody complexes were collected by the addition of Protein A–Sepharose (Pharmacia) and the immunocomplexes washed three times with 400  $\mu\text{l}$  RIPA buffer (10 mM Tris (pH 8.0), 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate). Immunoprecipitated complexes were resolved by SDS–PAGE on 10% gels which were then fixed in 30% methanol, 10% acetic acid, dried and autoradiographed. Gel mobility shift assays (20  $\mu\text{l}$ ) contained 10 mM Tris (pH 8.0), 40 mM KCl, 0.05% Nonidet P-40, 6% glycerol, 1 mM DTT, 0.2  $\mu\text{g}$  of poly(dI–dC) and 2.5  $\mu\text{l}$  each of *in vitro* synthesized RAR and RXR proteins. The total amount of reticulocyte lysate was maintained constant in each reaction (5  $\mu\text{l}$ ) through the addition of unprogrammed lysate. Where indicated, preimmune serum or polyclonal rabbit antisera prepared against bacterially expressed RAR $\alpha$  or an RXR $\alpha$  peptide (amino acids 214–229) were included. After a 10 min incubation on ice 1 ng  $^{32}\text{P}$ -labelled oligonucleotide was added and the incubation continued for an additional 10 min. DNA–protein complexes were resolved on a 4% polyacrylamide gel in 0.5 $\times$  TBE (1 $\times$  TBE is 90 mM Tris, 90 mM boric acid, 2 mM EDTA). Gels were dried and autoradiographed at  $-70^\circ$ . Gel mobility shift assays using COS cell-expressed receptors were as described<sup>11</sup>, using 5  $\mu\text{g}$  whole cell extracts prepared from COS cells transfected with 10  $\mu\text{g}$  of either pCMX–hRAR $\alpha$ , pCMX–hRXR $\alpha$ , or both expression plasmids.





Similar experiments in which RAR was replaced with radio-labelled glucocorticoid receptor (GR) failed to reveal RXR-GR interactions, demonstrating the specificity of the RAR-RXR interaction (Fig. 2a, lanes 5 and 6). Consistent with our transfection data indicating the importance of the C terminus of RAR in mediating RAR-RXR interactions, a truncated RAR protein (amino acids 155-462), lacking the amino terminus and DNA binding domain of the protein, was also efficiently coprecipitated with RXR (Fig. 2a, lanes 3 and 4). Thus, the C terminus of RAR, containing the dimerization domain, is sufficient for forming a stable solution complex with RXR.

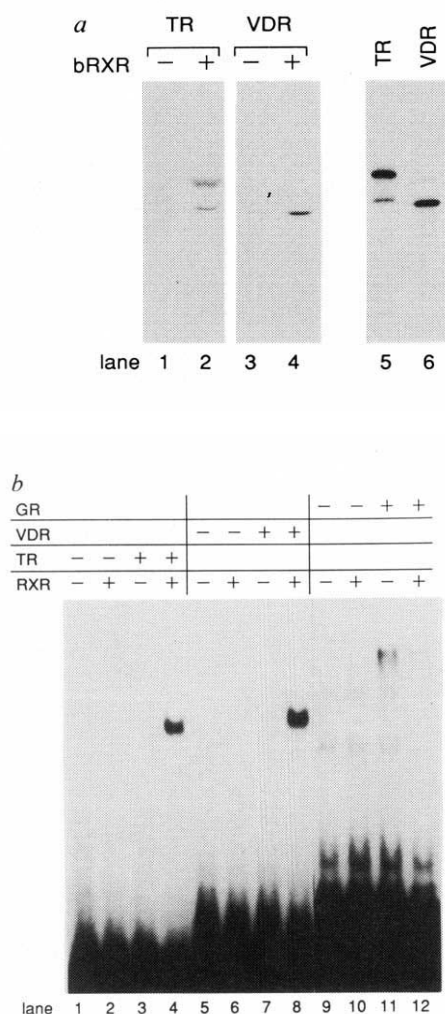


FIG. 3 Direct interactions between RXR and TR or VDR in the absence or presence of DNA. **a**, RXR forms complexes with either TR or VDR in solution. Immunoprecipitation reactions were done with RXR-specific antiserum and *in vitro* synthesized,  $^{35}\text{S}$ -methionine-labelled TR $\beta$  (lanes 1 and 2) or VDR (lanes 3 and 4) in either the absence (lanes 1, 3) or presence (lanes 2, 4) of bacterially expressed RXR (bRXR). Vitamin D $_3$  ( $1 \times 10^{-7}$  M) was included in reactions containing the VDR. *In vitro* synthesized TR and VDR proteins not subjected to immunoprecipitation are shown in lanes 5 and 6. **b**, RXR interacts cooperatively with TR and VDR in DNA binding. Gel mobility shift assays were done using *in vitro* synthesized RXR, TR, VDR and GR as indicated and  $^{32}\text{P}$ -labelled oligonucleotides encoding Moloney leukaemia virus LTR TRE $^{11}$  (lanes 1-4), the mouse osteopontin VDRE $^{11}$  (lanes 5-8), or the palindromic GRE $^{20}$  (lanes 9-12).

**METHODS.** TR $\beta$  and VDR RNA was prepared and subsequently translated in rabbit reticulocyte lysates as directed by the supplier (Promega). Immunoprecipitation and gel mobility shift assays are described in Fig. 2 legend.

These results led us to examine the properties of the RAR-RXR complex when associated with DNA. Gel mobility shift experiments were done using *in vitro* synthesized RAR and RXR and a radiolabelled oligonucleotide encoding the CRBP-II-RXRE. RAR synthesized *in vitro* bound with very low affinity to the CRBP-II-RXRE (Fig. 2b, lane 3). But the affinity of binding of RAR to CRBP-II-RXRE was greatly stimulated by the addition of *in vitro* synthesized RXR (Fig. 2b, lane 4). *In vitro* synthesized RXR alone had no detectable binding activity (Fig. 2b, lane 2). Inclusion of polyclonal antisera prepared against either RAR or RXR in the reaction resulted in complexes with reduced mobility (Fig. 2b, lanes 5 and 6), indicating that both RAR and RXR were present in the complex. Thus, the RAR-RXR complex is capable of interacting with the CRBP-II-RXRE with an affinity much higher than either receptor alone.

The specificity of the RAR-RXR interaction with DNA was next examined using unlabelled oligonucleotides as competitor. Oligonucleotides containing the CRBP-II-RXRE competed efficiently for RAR-RXR complex binding at a 10-fold molar excess (Fig. 2c, lane 2), whereas oligonucleotides containing an unrelated glucocorticoid response element (GRE) failed to compete when used at a 40-fold molar excess relative to the radiolabelled CRBP-II-RXRE (Fig. 2c, lane 7). Oligonucleotides containing the RARE of the RAR $\beta$  promoter ( $\beta$ RARE) $^{15,16}$  also competed efficiently for RAR-RXR binding to the CRBP-II-RXRE (Fig. 2c, lanes 4 and 5).

To investigate further this interaction of the RAR-RXR complex with the  $\beta$ RARE, oligonucleotides containing the  $\beta$ RARE were labelled and used as probe in a gel mobility shift assay. As in the case of the CRBP-II-RXRE, both *in vitro* synthesized RAR and RXR were required for high-affinity DNA-protein interactions with the  $\beta$ RARE (Fig. 2d, lanes 2-4). Similar results indicating a requirement for the presence of both RAR and RXR for formation of a high-affinity DNA-protein complex on the  $\beta$ RARE were obtained using whole-cell extracts prepared from COS cells which had been transfected with either RAR, RXR or RAR and RXR (Fig. 2e). Taken together, these results demonstrate that RXR dramatically stimulates the binding affinity of RAR to a strong RARE and that the RAR-RXR complex is likely to be present *in vivo*.

A functional relationship among the vitamin D receptor (VDR), thyroid hormone receptor (TR) and RAR has recently been described in which these receptors bind and activate through tandem repeats of consensus AGGTCA spaced by 3, 4 and 5 nucleotides, respectively (3-4-5 rule) $^{11}$ . Like the RAR, accessory factors present in nuclear extracts are necessary for high-affinity binding of the TR and VDR to their cognate response elements $^{6,9}$ . The relatively high degree of amino-acid conservation in the C termini of these nuclear receptors suggested that RXR might functionally interact with TR and VDR. Indeed, in immunoprecipitation experiments, *in vitro* synthesized TR and VDR coprecipitate with bacterially expressed RXR (Fig. 3a, lanes 1-4). The interactions of these receptors with RXR were also manifest at the level of DNA binding: *in vitro* synthesized RXR greatly stimulates TR and VDR binding to the Moloney leukaemia virus long terminal repeat thyroid hormone response element (MLV-LTR TRE) and osteopontin vitamin D response element (VDRE), respectively (Fig. 3b, lanes 1-8). The ability of RXR to stimulate the binding of nuclear receptors was not a general phenomenon, however, as RXR failed to increase GR binding to a GRE (Fig. 3b, compare lanes 11 and 12). Taken together, these data strongly suggest a central role for RXR in modulating the hormonal responses conferred through the RAR, TR and VDR.

The formation of RXR complexes with RAR, VDR and TR in which the complex displays new DNA-binding properties relative to the individual homodimers is reminiscent of interactions reported between the Jun and Fos families of proteins $^{17}$  as well as between members of the HLH family of transcription

factors such as MyoD and E12/47 (ref. 18). Through the formation of heterodimers, small families of structurally related proteins can yield large numbers of transcription factors with distinct functional properties. Two additional isoforms of RXR (RXR $\beta$  and RXR $\gamma$ ) have been recently identified (ref. 19, and D.J.M. and R.E., unpublished observations). Thus, the interaction of multiple RXR isoforms with additional nuclear receptors responsive to a diverse array of ligands is likely to have a critical role creating the high degree of diversity and specificity necessary to regulate the battery of hormone responsive genes.

Why the VDR, TR and RAR interact with a common partner is not yet clear, particularly as vitamin D<sub>3</sub> and thyroid hormone actions are not apparently retinoic acid-dependent. Further adding to the puzzle is the observation that RXR can activate through the CRBP11-RXRE in the absence of VDR, TR and RAR<sup>10</sup>, suggesting a role for other nuclear factors in this process. It is clear that characterization of the RXR family, its patterns of expression, and the nature of the RXR ligand will be essential to better understanding the complex molecular nature of hormonal signalling. □

Received 11 December; accepted 23 December 1991.

- Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. *Nature* **345**, 224–229 (1990).
- Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624–629 (1987).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444–450 (1987).
- Glass, C. K., Devary, O. V. & Rosenfeld, M. G. *Cell* **63**, 729–738 (1990).
- Rottman, J. N. *et al. Molec. cell. Biol.* **11**, 3814–3820 (1991).
- Liao, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9751–9755 (1990).
- Murray, M. B. & Towle, H. C. *Molec. Endocr.* **3**, 1434–1442 (1989).
- Burnside, J., Darling, D. S. & Chin, W. W. *J. biol. Chem.* **265**, 2500–2504 (1990).
- Lazar, M. A. & Berrodrin, T. J. *Molec. Endocr.* **4**, 1627–1635 (1990).
- Mangelsdorf, D. J. *et al. Cell* **66**, 555–561 (1991).
- Umesono, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. *Cell* **65**, 1255–1266 (1991).
- Fawell, S. E., Lees, J. A., White, R. & Parker, M. G. *Cell* **60**, 953–963 (1990).
- Forman, M. B. & Samuels, H. H. *Molec. Endocr.* **4**, 1293–1301 (1990).
- Glass, C. K., Lipkin, S. M., Devary, O. V. & Rosenfeld, M. G. *Cell* **59**, 697–708 (1989).

- Sucov, H. M., Murakami, K. K. & Evans, R. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5392–5398 (1990).
- de The, H. *et al. Nature* **343**, 177–180 (1990).
- Ransone, L. J. & Verma, I. M. *A. Rev. Cell Biol.* **6**, 539–557 (1990).
- Murre, C. *et al. Cell* **58**, 537–544 (1989).
- Hamada, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 8289–8293 (1989).
- Umesono, K. & Evans, R. M. *Cell* **57**, 1139–1146 (1989).
- Schule, *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 6092–6096 (1991).

ACKNOWLEDGEMENTS. We thank P. Rangarajan for RAR mutant expression constructs RS-Δ81–153 and RS-Δ203–360, R. Heyman for RAR mutant expression construct RS-185\*, J. Dyck for antisera prepared against RAR $\alpha$  and RXR $\alpha$  and H. Sucov, M. McKeown and T. Perlmann for discussions and critically reading the manuscript and E. Stevens for manuscript preparation. S.A.K. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. K.U. and D.J.M. are Research Associates and R.M.E. is an Investigator of the Howard Hughes Medical Institute at the Salk Institute for Biological Studies. This work was supported by the NIH and the Mathers Foundation.

## Control of DNA synthesis genes in fission yeast by the cell-cycle gene *cdc10*<sup>+</sup>

Noel F. Lowndes\*, Christopher J. McInerney†, Anthony L. Johnson\*, Peter A. Fantes† & Leland H. Johnston\*‡

\* Laboratory of Yeast Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

† Institute of Cell and Molecular Biology, University of Edinburgh, West Mains Road, Edinburgh EH9, UK

IN the budding yeast *Saccharomyces cerevisiae*, cell-cycle control over DNA synthesis occurs partly through the coordinate expression in late G1 phase of many, if not all, of the genes required for DNA synthesis<sup>1–3</sup>. A *cis*-acting hexamer element ACGCGT (an *Mlu*I restriction site) is responsible for coordinating transcriptional regulation of these genes at the G1/S phase boundary<sup>1,4</sup> and we have identified a binding activity, DSC1, that recognizes these sequences in a cell-cycle-dependent manner<sup>1</sup>. In the distantly related fission yeast<sup>5</sup> *Schizosaccharomyces pombe*, only one of the known DNA synthesis genes, *cdc22*<sup>+</sup>, which encodes a subunit of ribonucleotide reductase, is periodically expressed in late G1 (ref. 6). The promoter region of *cdc22*<sup>+</sup> has two *Mlu*I sites and five related sequences, suggesting that similar controls over DNA synthesis genes could occur in fission yeast. We report here a binding activity in fission yeast that is very similar to DSC1 in budding yeast. We also show that the fission yeast *cdc10*<sup>+</sup> gene product, which is required for Start and entry into S phase<sup>7,8</sup>, is a component of this binding activity.

We initially established that *Mlu*I motifs (Figs 1a and 2a) are physiologically significant in fission yeast. A plasmid containing three synthetic *Mlu*I sequences linked to the *Escherichia coli lacZ* gene was transformed into a wild-type strain and the cell-cycle expression of the *lacZ* transcript was examined in a synchronous culture (Fig. 1b). *LacZ* is expressed periodically. In the first cycle, expression of *lacZ* is slightly perturbed as it does not coincide precisely with *cdc22*<sup>+</sup> expression, which is

similar to our results with budding yeast<sup>1</sup>. But in the second cycle and even at the beginnings of a third cycle, *lacZ* and *cdc22*<sup>+</sup> are expressed in the same intervals of the cell cycle. Thus *Mlu*I sequences can confer expression in late G1 to a heterologous gene in *S. pombe*, as they do in budding yeast.

To detect a binding activity in *S. pombe* that recognizes *Mlu*I motifs, we used a probe containing three tandem *Mlu*I site repeats in gel retardation assays. With this probe we detect a binding activity in fission yeast which, like the budding yeast DSC1 activity, produces a retarded complex of very low mobility (Fig. 2b, lane 1). Furthermore, the specificity of the fission yeast activity is essentially identical to that of DSC1 (Fig. 2b), hence it will be referred to as 'DSC1-like'. In both cases the uppermost retarded complex is specifically competed away by multimeric copies of the *Mlu*I sequence, ACGCGT (Fig. 2b, lanes 2, 3 and 17, 18), but not by multimeric copies of a mutated version of this sequence, ACTaGT (Fig. 2b, lanes 4, 5 and 19, 20), nor by multimeric copies of a similar oligonucleotide containing an unrelated hexamer, ATGATT (Fig. 2b, lanes 14, 15 and 29, 30). Therefore, both binding activities recognize the ACGCGT core itself and not other sequences found on the oligonucleotide used as a probe (see legend to Fig. 2).

The 132-base pair (bp) and 134-bp fragments from the *cdc22*<sup>+</sup> promoter containing four and two *Mlu*I elements (Figs 1a and 2a), respectively, each compete successfully for binding of both the budding yeast DSC1 activity and of the fission yeast DSC1-like activity (Fig. 2b; compare lanes 8–11 and 23–26). Thus both activities recognize *Mlu*I motifs in their normal sequence context as they occur in the *cdc22*<sup>+</sup> promoter. Similarly, the fission yeast binding activity recognized the *Mlu*I sites in the promoter region of the periodically expressed DNA synthesis gene *CDC9* (DNA ligase) from budding yeast<sup>10</sup>. A 55-bp fragment from this promoter<sup>1</sup>, which contains a single *Mlu*I site and a near-match (Fig. 2a), competes successfully for binding of both DSC1 and the DSC1-like activity from fission yeast. (Fig. 2b, lanes 12, 13 and 27, 28). In contrast, a 166-bp fragment from the *cdc22*<sup>+</sup> upstream containing only a single *Mlu*I near-match (Fig. 2a) did not compete for binding of either the fission or the budding yeast activities (Fig. 2b, lanes 6, 7 and 21, 22). These experiments have been repeated using the 132-bp fragment from the *cdc22*<sup>+</sup> upstream as a labelled substrate, which gives essentially identical results (data not shown). In conclusion, the identical competition profiles for both binding activities, particularly when using

‡ To whom correspondence should be addressed.



**a**  
 GAAATTAGTAGTTCAATCTCATAGAGCAGGTTGGTAGTCGGGTTGGACCGCGTGT  
 TTAATTTATGTAACAGTTCGCGTTCGCGTTGCAATTGAGACGCGTAATAAATATTT  
 AATTTATTACATTCACTGTAACAGAGTATTTATAAACACITTTTTTATGTTTAAT  
 AAAAGATAAATGTAACAGTTGAATGTATGTATCAGGTCAGACCACTTCAACATG  
 TTTAAATCGCGTTTTTTTAAAAAATATTTTTTAAATTTTAAAAAGTCGGACTTATTT  
 AGCGGAACTTTGATGTTTCAGAAAGTAAAAAGATAAATCTATTTAGCAAGTCTTAATT  
 AACGCTCTTTTATAGATATAGTAGAGCTACAAAATGATCCGGTTTCCACTCTTAGCTT  
 TATTTACATTGATCAACATGACTTAAAGTTCGGATGACCGGACGGGCCACATCTCA  
 ACCGGTTTAAATTAATAGCGTGACCGGTCGAACGCTTTTTCATTACTATAAATAT  
 TCCCGGTATTTACCACCTTACAACTAGAACAAACACGATG

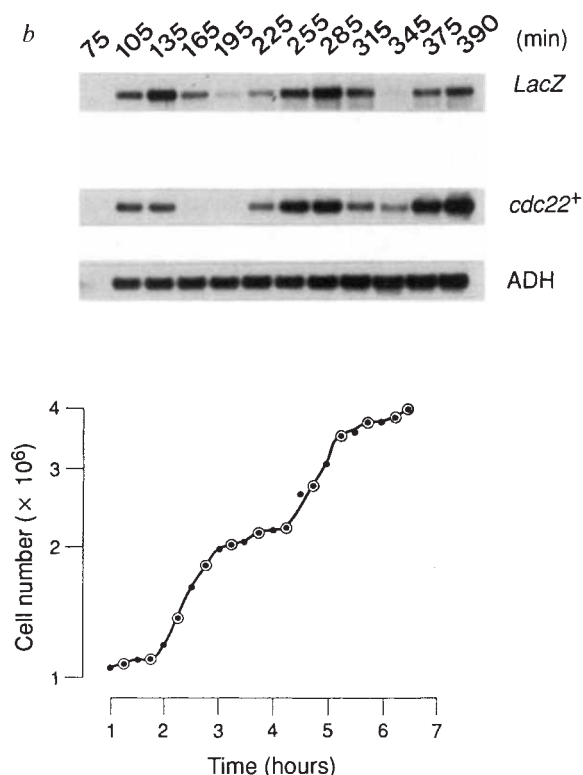


FIG. 1 **a**, The DNA sequence of the *S. pombe* *cdc22*<sup>+</sup> promoter region. The likely ATG at the start of translation is shown in bold and the asterisk marks the start of transcription (C.J.McI. and P.A.F., manuscript in preparation). The two *Mlu*I sequences, ACGCGT, are italicized and underlined and the near-match sequences that are likely to be significant<sup>2</sup> are also underlined. **b**, *Mlu*I motifs confer periodic expression on a reporter gene in *S. pombe*. Northern blot analysis of RNA was carried out on cells synchronized by elutriation and containing a test plasmid with three ACGCGT sequences linked to *lacZ*. The transcript levels of *lacZ* and *cdc22*<sup>+</sup> are shown together with alcohol dehydrogenase (ADH), which acted as a control to show that the gel loading was even. The times in minutes at which samples were taken are shown above each lane. Note that the RNA in lane 1 was degraded during preparation. The lower panel shows the increase in cell numbers during the experiment.

**METHODS.** The budding yeast test plasmid pLGΔ178.3M<sup>1</sup> containing three repeats of the *Mlu*I sequence makes use of the *S. cerevisiae* cytochrome *c* minimal promoter which would be expected to function in *S. pombe*<sup>9</sup>. As it also carries the *S. cerevisiae* *URA3* gene which can be selected in a *S. pombe* *ura4* mutant, the plasmid was adapted for use in fission yeast. The *S. cerevisiae* 2  $\mu$ m origin of replication was removed by digestion with *Eco*RI and replaced by ARS6, a *S. pombe* chromosomal origin of replication. The resulting vector, pSPΔ178.3M, was transformed into *S. pombe* h<sup>-</sup> *ura4* and synchronized by elutriation<sup>6</sup>. Cell numbers were monitored using a Coulter counter and samples were removed at intervals (circled points in the lower panel). RNA extraction, northern blotting and probing have been described<sup>10</sup>.

promoter fragments from periodically expressed DNA synthesis genes from these two highly divergent organisms<sup>5</sup>, strongly argues that the two binding activities are functionally conserved.

Mutants defective in the DSC1 and DSC1-like binding activities would greatly facilitate their further characterization. As these activities control genes essential for DNA synthesis, such mutants might themselves be cell-cycle mutants with execution points in late G1. In the fission yeast, only *cdc2*<sup>+</sup> and *cdc10*<sup>+</sup> are known to act in late G1 and prevent DNA synthesis<sup>7,8,11</sup>. Significantly, the *cdc10*<sup>+</sup> gene product is homologous with the budding yeast SWI4 and SWI6 gene products<sup>12,13</sup>, which are believed to be transcriptional activators<sup>13,14</sup>. Furthermore, they actually regulate the cell-cycle specific expression of the HO endonuclease (involved in mating type switching) which is periodically transcribed in late G1 (refs. <sup>13-16</sup>). Thus the *cdc10*<sup>+</sup> gene may have a similar function. Consistent with this, *cdc10*<sup>-</sup> mutants blocked at the restrictive temperature have substantially lower *cdc22*<sup>+</sup> transcript levels than in wild-type cells (M.J.F.S. and P.A.F., unpublished results). We have now found that *cdc10*<sup>-</sup> mutants are also defective in their DSC1-like activity (Fig. 3). When *cdc10*-129 mutant cells are shifted to the restrictive temperature there is a rapid loss of DSC1-like binding activity which is not observed in similarly treated wild-type cells (Fig. 3a; compare lanes 1-6 with 7-12). This loss was evident after only 5 min at 37 °C and there was no binding activity after 30 min, which is well within one cell cycle, so the loss cannot be a secondary consequence of a cell-cycle block.

To confirm this result, we prepared crude extracts from cells containing one of three *cdc10*<sup>-</sup> mutant alleles grown at the permissive temperature. Extracts were incubated at 28 °C for 0, 5, 10 and 15 min before assaying their binding activities (Fig. 3b). In one allele, *cdc10*-C4, no DSC1-like binding activity was detectable even before the incubation at 28 °C (Fig. 3b, lanes 13-16), which presumably indicates that the protein product is extremely unstable. The other two alleles, *cdc10*-129 and *cdc10*-K28, produced much reduced DSC1-like activity, which was rapidly inactivated at 28 °C *in vitro* (Fig. 3b, lanes 5-8 and 9-12). This loss of activity is not simply due to inhibition of binding in the mutant cells, because mixing extracts from mutant and wild-type cells has no effect on wild-type activity (Fig. 3c). Furthermore, three revertants of *cdc10*-129 that are able to grow at 37 °C contained binding activity that was insensitive to the temperature shift (Fig. 3d, lanes 1-6). Finally, introduction of a plasmid carrying the *cdc10*<sup>+</sup> gene into the *cdc10*-129 mutant strain fully restores the binding activity at the restrictive temperature (Fig. 3d, lanes 7 and 8). Thus the temperature-dependent loss of DSC1-like activity in *cdc10*<sup>-</sup> cells seems to be due to a temperature-sensitive gene product and suggests strongly that the *cdc10*<sup>+</sup> protein is directly involved with the binding activity.

To determine whether the *cdc10*<sup>+</sup> gene product itself is a component of the fission yeast DSC1-like activity, we used anti-*cdc10*<sup>+</sup> polyclonal serum, which when added to a binding reaction caused a super-shift of the DSC1-like retarded complex (Fig. 4a, lanes 3-7), an effect not seen with preimmune serum (Fig. 4a, lane 2). When high concentrations of the anti-*cdc10*<sup>+</sup> serum were used, the super-shifted complex did not even enter a polyacrylamide gel (Fig. 4a, lanes 3, 4), but in agarose its electrophoretic mobility increased when the serum was diluted (Fig. 4b, lanes 3-7), presumably because fewer of the available epitopes are bound. This shows that the *cdc10*<sup>+</sup> gene product is itself part of the fission yeast DSC1-like activity.

Given the homology of *cdc10*<sup>+</sup> with SWI4 and SWI6, the *cdc10*<sup>+</sup> protein is very likely to be a transcription factor. Also, the occurrence of the 'SWI6-*cdc10*<sup>+</sup>' motif (also known as the 'ankyrin' repeat) in these three proteins<sup>12,13,17</sup> suggests that they may be analogous to the  $\beta$ -subunit of the recently characterized heteromeric GABP transcription factor<sup>18</sup>. The *cdc10*<sup>+</sup> protein is probably not the only component of the fission yeast *Mlu*I binding activity as several of the retarded species shown in

FIG. 2 Gel retardation analysis showing binding of specific proteins from *S. pombe* and *S. cerevisiae* to the ACGCGT sequence element. *a*, The promoter regions of the periodically expressed *cdc22<sup>+</sup>* gene from *S. pombe* and *CDC9* gene from *S. cerevisiae* and the relative positions of the DNA fragments used in competition experiments. The scale is in base pairs and numbered relative to the ATG codon. Filled boxes represent perfect matches to the ACGCGT sequence element, whereas open boxes represent near-matches where either the A or the T are altered from the *MluI* consensus<sup>2</sup>. The wavy arrows represent the position of the transcripts for both genes. The 132 bp, 166 bp and 134 bp DNA fragments were generated by the polymerase chain reaction, whereas the 55-bp fragment was generated by digestion with *MnII* and *AluI* restriction endonucleases. *b*, Clarified crude extracts were prepared from the *S. cerevisiae* strain CG378 and from *S. pombe* wild-type cells and 20  $\mu$ g or 5  $\mu$ g aliquots, respectively, were used with a 73-bp *AluI/Sau3A* fragment from pLG $\Delta$ 178.3M as previously described<sup>1</sup>. The competitor DNAs used were as follows (in each case a 10-fold and a 100-fold molar excess was used): lanes 1 and 16, no specific competitor; lanes 2, 3 and 17, 18 a tetrameric form of the *MluI* site oligonucleotide, 5'-TCGAGACGCGTC-3'; lanes 4, 5 and 19, 20 a tetrameric form of the mutated *MluI* site oligonucleotide, 5'-TCGAGACtaGTC-3'; lanes 6, 7 and 21, 22, the 166-bp fragment from the *cdc22<sup>+</sup>* upstream; lanes 8, 9 and 23, 24, the 132-bp fragment; lanes 10, 11 and 25, 26, the 134-bp fragment; lanes 12, 13 and 27, 28, the 55-bp fragment from the *CDC9* upstream; lanes 14, 15 and 29, 30, multimeric form of an oligonucleotide containing an unrelated hexamer, 5'-TCGAGATGATTC-3'. The multimeric forms of the competitor oligonucleotides were generated by ligation and either total ligation products were used or the tetrameric form was purified by gel electrophoresis. The small arrowhead indicates the specific DSC1 and DSC1-like retarded products, the larger arrowhead indicates the unbound probe. Brackets indicate specific retarded species, which may correspond to proteolytic products of the uppermost retarded complexes as gradual loss of these due to proteolysis is accompanied by a concomitant increase in the faster-migrating species (unpublished observations). The fission yeast activity is more readily detectable in crude extracts

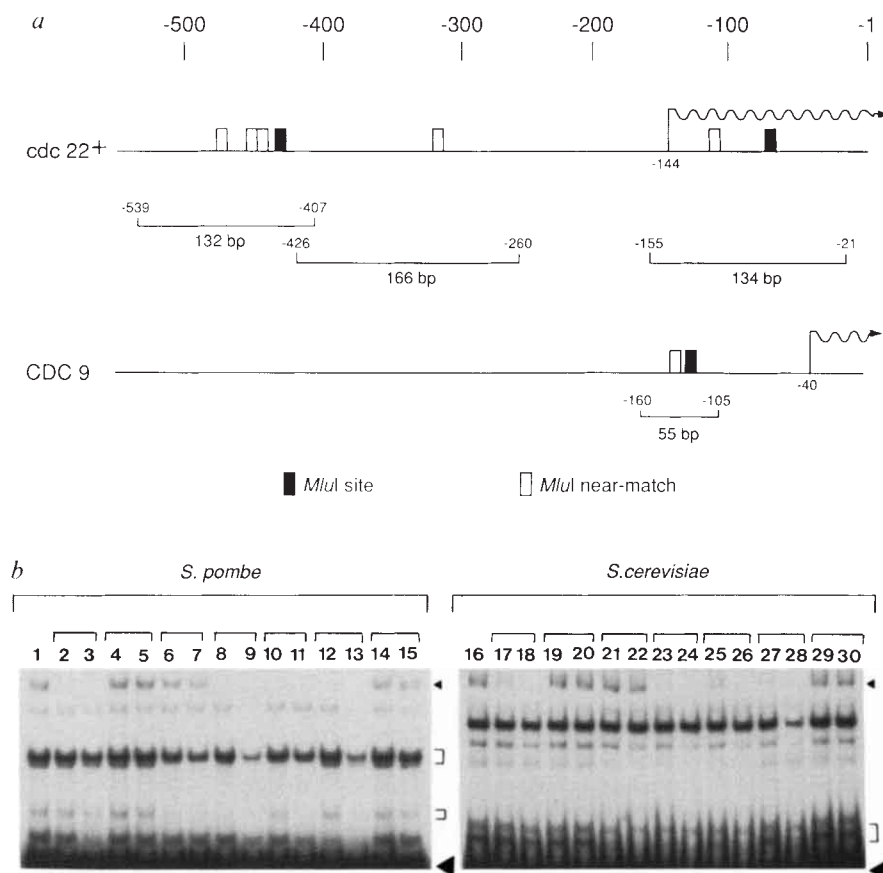


Fig. 2b are competed out by *cdc22<sup>+</sup>* and *CDC9* upstream fragments. Whereas the faster migrating species are partly due to proteolysis of the uppermost complex (see legend to Fig. 2), it could be significant that only this uppermost species is absent in *cdc10<sup>-</sup>* mutants and it alone is affected by the *cdc10<sup>+</sup>* antiserum.

The striking similarity between the mobility and specificity of the budding and fission yeast DSC1 activities argues that these binding activities have been functionally conserved in both organisms. But much of the molecular detail has to be established, for instance the similarity between the proteins in the retarded complexes (the *cdc10<sup>+</sup>* antiserum we used did not bind to the budding yeast complex) and the regulation of the binding activities. As the *cdc10<sup>+</sup>* gene product is a phosphoprotein<sup>19</sup>, the *cdc2<sup>+</sup>* protein kinase might be involved in its phosphorylation. When *cdc10<sup>+</sup>* is present on a high-copy-number plasmid, the level of the DSC1-like activity is not increased significantly above the wild-type level (Fig. 3d, lanes 7, 8), which is consistent with some form of post-translational control. Alternatively,

then the budding yeast activity, so to give comparable autoradiographic times and reveal all the retarded species, only 5  $\mu$ g *S. pombe* crude extract was used in this experiment.

**METHODS.** All binding reactions were performed with 0.5 ng  $\gamma$ -<sup>32</sup>P-5' end-labelled 73-bp *AluI/Sau3A* fragments at  $\sim 1 \times 10^8$  c.p.m.  $\mu$ g<sup>-1</sup>. The preparation of clarified crude extracts, protein estimations, binding buffer and gel running conditions were as before<sup>1</sup>, except that electrophoresis was through nondenaturing 4% (40:1) polyacrylamide gels.

assembly of the complex could be limited by a component other than the *cdc10<sup>+</sup>* protein.

It is nevertheless clear that the mechanisms for regulating DNA synthesis genes are similar in both organisms and, therefore, that this particular aspect of DNA synthesis control is conserved. The system is more limited in fission yeast, for example its DNA ligase gene *cdc17<sup>+</sup>* is not cell-cycle regulated, unlike *CDC9* in budding yeast<sup>10</sup>. Indeed, of the DNA synthesis genes known in the fission yeast only ribonucleotide reductase (*cdc22<sup>+</sup>*) seems to be cell-cycle regulated<sup>6</sup>. Interestingly, this enzyme is cell-cycle regulated in many organisms (for a review, see ref. 20), which suggests that it could be rate-limiting and therefore have a widespread role in controlling DNA synthesis. Whether it is regulated in these other organisms by a system of *MluI* sites, or related sequences, together with a DSC1-like binding activity is not yet clear. However, budding and fission yeasts are widely divergent evolutionarily<sup>5</sup> so that systems conserved in both organisms are frequently widespread in nature. □



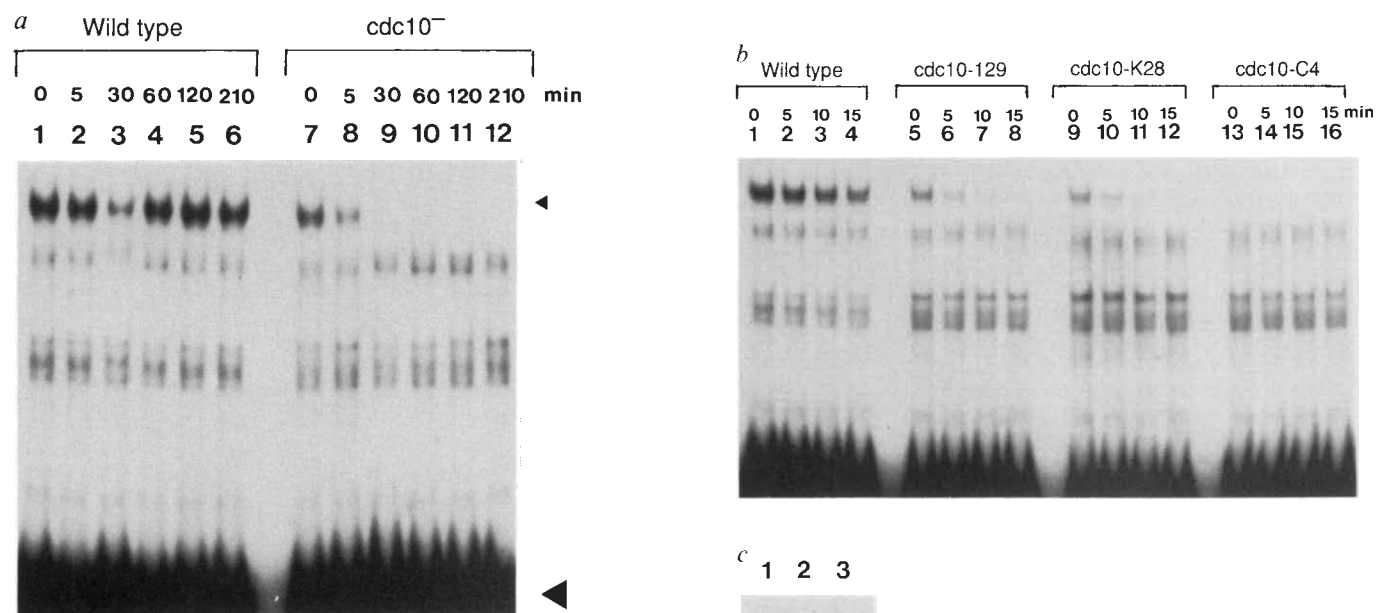


FIG. 3 Gel retardation analysis of the 'DSC1-like' activity in *S. pombe cdc10*<sup>-</sup> mutants. **a**, *In vivo* temperature shift. Mid-log cultures of *S. pombe* wild type (lanes 1–6) and *cdc10*-129 cells (lanes 7–12), grown at 25 °C, were shifted to 37 °C and samples taken at 0, 5, 30, 60, 120 and 210 min. Clarified crude extracts were prepared and 10 µg wild type or 20 µg *cdc10*-129 aliquots were assayed for DSC1-like activity. The small arrowhead indicates the specific DSC1-like retarded product and the larger arrowhead indicates the unbound probe. **b**, *In vitro* temperature shifts. Clarified crude extracts were prepared from wild type (lanes 1–4), *cdc10*-129 (lanes 5–8), *cdc10*-K28 (lanes 9–12) and *cdc10*-C4 (lanes 13–16) cells grown at 25 °C to mid-log phase. These clarified crude extracts were then incubated at 28 °C for 0, 5, 10 and 15 min respectively, as indicated, immediately before use in binding assays. Note that 10 µg wild-type extract and 20 µg *cdc10*<sup>-</sup> mutant extract was used for each time point. **c**, Mixing experiment. Clarified crude extracts from wild-type cells (10 µg; lane 1) or *cdc10*-129 mutant cells held at the restrictive temperature for 1 h (20 µg; lane 2) were assayed either individually or mixed together (10 µg wild type + 20 µg *cdc10*-129; lane 3). **d**, Analysis of revertants and plasmid rescue of the mutant phenotype. Three spontaneously occurring intragenic revertants (confirmed by genetic analysis) of the *cdc10*-129 allele were isolated, grown to mid-log phase at 25 °C, sampled (lanes 1, 3 and 5), shifted to 37 °C for 1 h and then sampled once more (lanes 2, 4 and 6). Lanes 1, 2, revertant 1; lanes 3, 4 revertant 2; lanes 5, 6 revertant 3. The *cdc10*-129 mutant strain was also

transformed with a plasmid carrying the wild type *cdc10*<sup>+</sup> gene, grown at 25 °C to mid-log phase, sampled (lane 7), then shifted to 37 °C for 1 h and sampled again (lane 8).

METHODS. All binding reactions were performed as described before<sup>1</sup> and in the legend to Fig. 2, using 10 µg clarified crude extract unless otherwise stated.

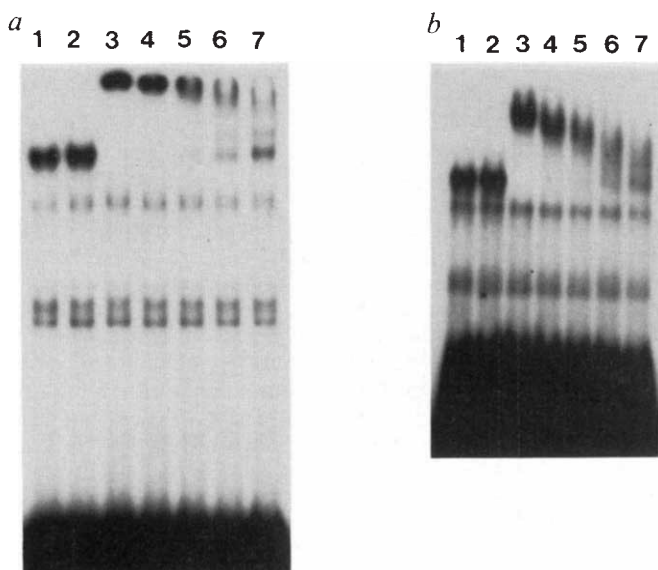


FIG. 4. The DSC1-like activity of *S. pombe* contains the *cdc10*<sup>+</sup> gene product. Binding reactions were allowed to equilibrate as before<sup>1</sup> (see also legend to Fig. 2) and then further additions were as follows: lane 1, binding buffer; lane 2, 1/12.5 dilution of pre-immune serum; lanes 3–7, a dilution series of anti-*cdc10*<sup>+</sup> serum, 1/12.5, 1/25, 1/50, 1/100, 1/200. Equal volumes of each reaction were then loaded onto either **a**, 4% (40:1) non-denaturing polyacrylamide or **b**, 2% Nusieve: 1% agarose gels.

METHODS. All binding reactions were performed as described previously<sup>1</sup> and in Fig. 2 legend, except that once equilibrated on ice a further 1 µl of either binding buffer or the appropriate dilution of serum was added and incubated for a further 5 min on ice before loading onto gels. Gels were run as before<sup>1</sup>. Both the preimmune and immune sera were from the same rabbit.

Received 2 August; accepted 7 November 1991.

- Lowndes, N. F., Johnson, A. L. & Johnston, L. H. *Nature* **350**, 247–250 (1991).
- Johnston, L. H., Lowndes, N. F., Johnson, A. L. & Sugino, A. *Cold Spring Harb. Symp. quant. Biol.* (in the press).
- Andrews, B. J. & Herskowitz, I. *J. biol. Chem.* **265**, 14057–14060 (1990).
- MacIntosh, E., Atkinson, T., Storms, R. K. & Smith, M. *Molec. cell. Biol.* **11**, 329–337 (1991).
- Sipiczki, M. in *Molecular Biology of the Fission Yeast* (eds Nasim, A., Young, P. & Johnson, B.) 431–452 (Academic, New York, 1989).
- Gordon, C. B. & Fantes, P. A. *EMBO J.* **5**, 2981–2986 (1986).
- Nurse, P., Thuriaux, P. & Nasmyth, K. *Molec. gen. Genet.* **146**, 167–178 (1976).
- Nurse, P. & Bissett, Y. *Nature* **292**, 558–560 (1981).
- Russell, P. R. *Nature* **301**, 167–169 (1983).
- White, J. H. M., Barker, D. G., Nurse, P. & Johnston, L. H. *EMBO J.* **5**, 1705–1709 (1986).
- Nasmyth, K. A. & Nurse, P. *Molec. gen. Genet.* **182**, 119–124 (1981).
- Breeden, L. & Nasmyth, K. *Nature* **329**, 651–654 (1987).
- Andrews, B. J. & Herskowitz, I. *Nature* **342**, 830–833 (1989).
- Andrews, B. J. & Herskowitz, I. *Cell* **57**, 21–29 (1989).
- Nasmyth, K. *Nature* **302**, 670–673 (1983).
- Breeden, L. & Nasmyth, K. *Cell* **48**, 389–397 (1987).
- Aves, S. J., Durkacz, B. W., Carr, A. & Nurse, P. *EMBO J.* **4**, 457–463 (1985).
- Tompson, C. C., Brown, T. A. & McKnight, S. L. *Science* **253**, 762–768 (1991).
- Simanis, V. & Nurse, P. *J. Cell. Sci.* **92**, 51–56 (1989).
- Fernandez-Sarabia, M. J. & Fantes, P. A. *Trends Genet.* **6**, 275–276 (1990).

ACKNOWLEDGEMENTS. We thank U. Deuschle, D. Beach and V. Simanis for *cdc10*<sup>+</sup> antiserum; J. Crezanor for *cdc10*-C4 and for elutriation; and M. J. Fernandez-Sarabia for unpublished information.

## Expression of members of the putative olfactory receptor gene family in mammalian germ cells

Marc Parmentier\*, Frédéric Libert\*, Stéphane Schurmans\*, Serge Schiffmann†, Anne Lefort\*, Dominique Eggerickx\*, Catherine Ledent\*, Catherine Mollereau‡, Catherine Gérard\*, Jason Perret\*, Anton Grootegoed§ & Gilbert Vassart\*\*

\*IRIBHN and \*\*Service de Génétique and †Laboratoire de Recherche sur les neuropeptides, Université Libre de Bruxelles, Campus Erasme, 808 route de Lennik, 1070 Bruxelles, Belgium

‡Laboratoire de Pharmacologie et de Toxicologie fondamentale du CNRS, 205 route de Narbonne, 31077 Toulouse, France

§Department of Endocrinology and Reproduction, Medical Faculty, Erasmus University Rotterdam, Rotterdam, The Netherlands

A SERIES of genomic and complementary DNA clones encoding new putative members of G protein-coupled receptors were isolated using homology cloning and low-stringency polymerase chain reaction<sup>1,2</sup>. Among the unidentified receptors ('orphan receptors'), a human genomic clone (HGMP07) was characterized by the presence of its transcripts in the testis and by its belonging to a large subfamily of genes sharing extensive sequence similarities. Sequence comparison demonstrated that this gene subfamily is the human counterpart of the putative rat olfactory receptors cloned recently<sup>3</sup>. Another 48 members of the family were cloned. Northern blotting further demonstrated the presence of olfactory receptor transcripts in germ cells. Our finding suggests that a common receptor gene family encodes olfactory receptors and sperm cell receptors that could be involved in chemotaxis during fertilization.

In the search for the thyrotropin receptor, we used low-stringency polymerase chain reaction (PCR)<sup>1</sup> to clone a series of new G protein-coupled receptor genes directly from genomic DNA (ref. 2). A couple of degenerate primers were devised on the basis of conserved sequences in the second, third, sixth and seventh transmembrane segments of known members of this gene family. Eleven new candidate receptor gene fragments were cloned in this way<sup>2</sup>. To target the ligand screening for these orphan receptors, northern blots of nine dog tissues were systematically probed for hybridization with each receptor sequence. One clone, HGMP07, detected transcripts in the testis only (Fig. 1a). A human genomic DNA library, and a human testis cDNA library were screened for full-length clones corre-

sponding to HGMP07. The result was the isolation of 16 different clones with strong similarities with HGMP07: eight genomic clones and nine testicular cDNAs, one receptor (HGMP07J) being isolated from both libraries. Genomic clones contained an open reading frame colinear with the cDNA sequences, demonstrating the absence of introns in the coding region. Sequence comparison was made with the steadily increasing number of G protein-coupled receptors<sup>4,5</sup>. In the recent cloning of putative rat olfactory receptors<sup>3</sup>, it seemed that these olfactory receptors constituted the rat counterpart of our subfamily of human sequences (Fig. 2). All landmarks of the putative olfactory receptors were found, both in the genomic and the testicular cDNA clones. Sequence conservation was maximal in the second, third, sixth and seventh transmembrane segments.

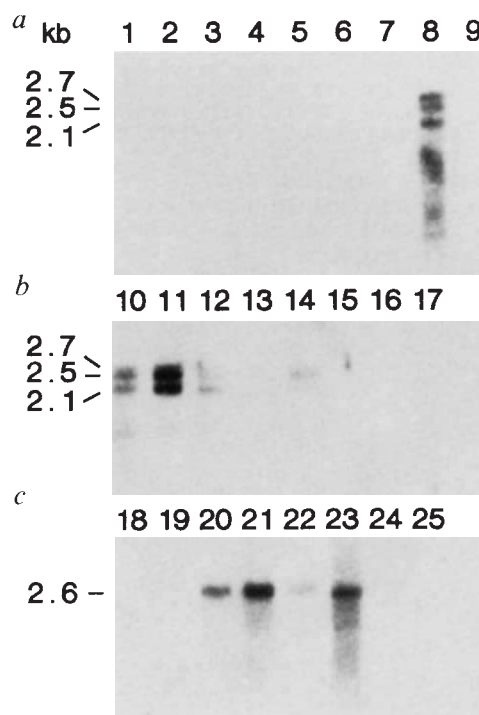


FIG. 1 Northern blotting showing the tissue and cellular distribution of transcripts from the putative olfactory receptor family. a, Northern blot of RNA extracted from nine dog tissues: stomach, spleen, kidney, liver, heart, lung, brain, testis and thyroid (lanes 1–9, respectively). The blot was hybridized with the original HGMP07 PCR fragment as a probe. b, Northern blot of RNA extracted from fractionated dog germ cells (cell preparations enriched in late elongated spermatids (lane 10), round spermatids and pachytene spermatocytes (lane 11), Sertoli cells (lane 12), and from dog ovary, testis, brain, olfactory mucosa and lung (lanes 13–17, respectively). The blot was hybridized with the dog DTMT clone (see text and Fig. 2). c, Northern blot of RNA (same preparations as in Fig. 1b) extracted from dog olfactory mucosa (lane 18), cell preparations enriched in late elongated spermatids (lane 19), round spermatids and pachytene spermatocytes (lane 20), Sertoli cells (lane 21), and from dog testis, ovary, lung and brain (lanes 22–25, respectively). The blot was hybridized with a human FSH receptor probe (HGMP09, ref. 2).

METHODS. Testicular tubules were prepared from dog testis by enzymatic digestion using collagenase, hyaluronidase, trypsin and DNase I (ref. 9). These tubules, containing Sertoli cells and germ cells (tubular wall and interstitial tissue was removed by enzyme treatment) were treated once more with trypsin and DNase I (ref. 9), followed by mechanical dispersion. Clumps of Sertoli cells (with attached germ cells including some spermatids) were pelleted at 20g (2 min). Pachytene spermatocytes and round spermatids (and a small number of single Sertoli cells) were pelleted at 60g (4 min) and the remaining round and elongating spermatids were pelleted at 600g (4 min). Poly(A)<sup>+</sup>RNA was prepared by classical caesium chloride centrifugation<sup>10</sup> or by using the Fastrack kit (Invitrogen). Northern blotting was as described<sup>11</sup>.





FIG. 2 Amino-acid sequences (single-letter code), deduced from the HGMP07I, DTMT and HGMP07J clones, compared with two rat olfactory clones described by Buck and Axel<sup>3</sup>. HGMP07I is a human genomic clone corresponding to the original HGMP07 probe obtained by PCR. DTMT represents the major transcript detected in dog sperm cells. HGMP07J is one of the genes expressed in human testis. Bars and roman numbers represent the putative transmembrane segments. Asterisks and dots below the alignment indicate

identical residues and conservative replacements, respectively. Putative sites for *N*-linked glycosylation are underlined. HGMP07J has two possible initiation sites, the second of which is aligned with the other putative receptors.

**METHODS.** Clones were isolated from a human testis cDNA library constructed in  $\lambda$ gt11 and from human and dog genomic libraries using standard procedures<sup>10</sup>. Sequencing was on an Applied Biosystem sequencer.

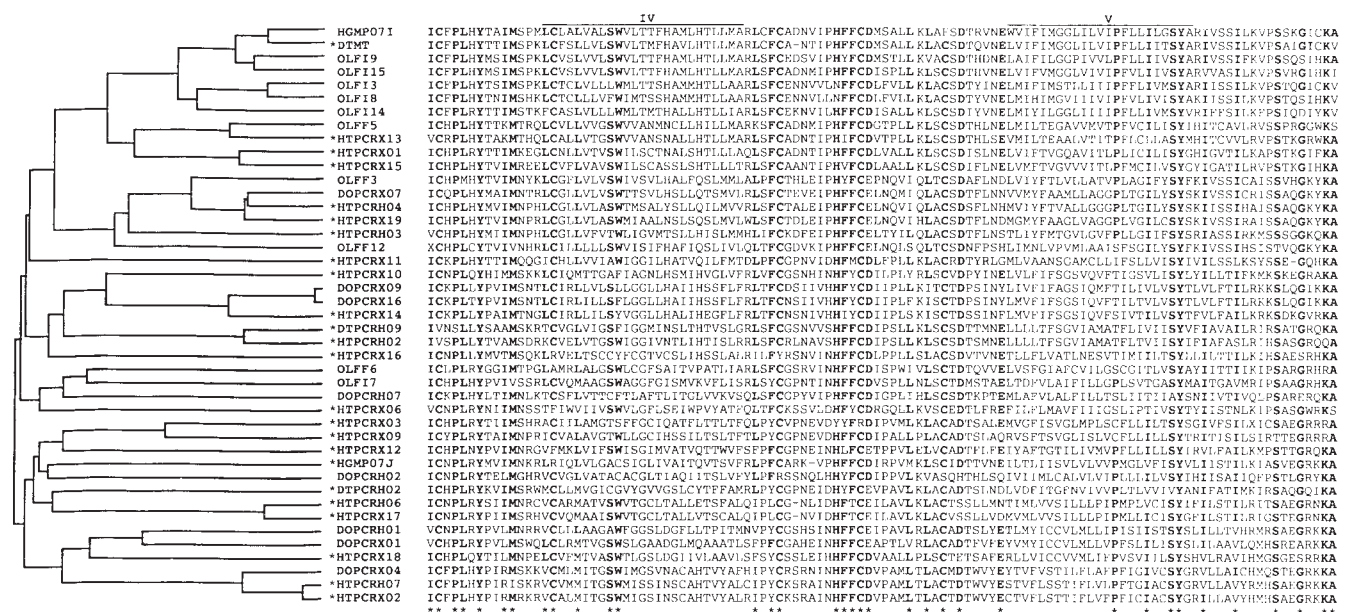


FIG. 3 Alignment of partial sequences from 43 members of the olfactory receptor family HGMP07I, DTMT and HGMP07J clones are described in Fig. 2. OLF clones are from Buck and Axel<sup>3</sup>. Other clones were obtained by low-stringency PCR. HTPCR, DTPCR and DOPCR clones were isolated from reverse transcripts of human testis, dog testis and dog olfactory mucosa, respectively. Bars and roman numbers represent the putative transmembrane segments. Asterisks below the alignment indicate positions where a specific amino acid is present in more than two-thirds of the sequences. The dendrogram represents the relative homology between the displayed sequences. Asterisks tag names of human and dog clones that

are expressed in testis.

**METHODS.** Low-stringency PCR was done essentially as described<sup>1-3</sup> using primers specific to the olfactory receptor family: ATGGC(ITA)(TC)-GA(TC)(CA)G(AT)(TC)GTIGC and A(GA)I(GC)(AT)UACIA(CT)I(GC)(AT)I(A)(GA)-(GA)TG(IG)(CA)(AT)I(GC)(C)(AG)CAIGT. The mRNA was treated with DNase I (10 U  $\mu\text{g}^{-1}$  RNA) before reverse transcription to avoid amplification from contaminating genomic DNA. The multiple sequence alignment and the resulting dendrogram was made with the Clustal software<sup>1,2</sup>. The dendrogram represents the relative sequence similarities between the members of the gene family.

All sequences displayed one conserved potential site for N-linked glycosylation in the short N-terminal extracellular domain.

Degenerate primers specific for the subfamily were designed and used in PCR experiments on dog and human messenger RNA that had previously been treated with DNase and reverse-transcribed. This led to the cloning of three partial clones from dog testis, 21 from human testis, and eight from dog olfactory mucosa (Fig. 3). Multiple copies of one clone (DTPCHRH01; Fig. 3) were obtained from the PCR reaction on dog testis mRNA, suggesting that it could represent the most abundant transcript of this receptor family. Northern blots with this probe confirmed this relative abundance. The full coding region corresponding to this PCR clone was subsequently isolated from a dog genomic library and named DTMT (Fig. 2). The alignment of most of the available partial sequences extending between transmembranes III and VI is shown in Fig. 3, together with a dendrogram representing the structural relationships between the members of the subfamily. As noted by Buck and Axel<sup>3</sup>, individual clones can be classified into subcategories. The percentage of identical residues between any two sequences ranges from 30 to over 90%. Expression in olfactory mucosa or testis cannot be correlated with the receptors belonging to specific subfamilies. The clone relatively highly expressed in dog testis (DTMT) shares 82% identity with OLF15, a rat cDNA cloned from the olfactory mucosa<sup>3</sup>. No data support a single receptor gene being expressed in both locations (see below). The subset of receptors described by Buck and Axel do not seem representative of all subfamilies. This supports the assumption that their estimated number of olfactory receptor genes (100–200 per haploid genome) represented a lower limit. The screening of a dog genomic library with a mixture of PCR clones allows us to estimate the size of the family to more than 400 genes, and probably this number is still an underestimate.

To investigate the presence of candidate olfactory receptor transcripts in the testis, additional northern blots prepared with RNA from testicular fractions enriched in different cell types were probed with DTMT (Fig. 1b). The results demonstrate a high expression in fractions enriched in spermatocytes and spermatids. The data do not give information about the onset of expression (pre- or postmeiotic). Expression of putative olfactory receptor transcripts was low in a fraction enriched in Sertoli cells, expressing a high level of FSH receptor mRNA (this receptor is unique to Sertoli cells; Fig. 1c). The DTMT transcript could hardly be detected in the dog olfactory mucosa. But considering the diversity of the gene family and the expected restriction of expression of individual receptors to a limited number of cells, this does not exclude expression of DTMT in olfactory mucosa.

The presence of transcripts belonging to the putative olfactory receptor gene family in germ cells does not necessarily imply that the corresponding proteins are produced. But it is tempting to speculate that they might encode receptors implicated in the chemotaxis of sperm cells during fertilization. Sperm chemotaxis, involving receptors of the guanylate cyclase family, has been clearly established in sea urchin<sup>6</sup>. Experimental evidence for the existence of a similar phenomenon in mammals has recently been provided<sup>7,8</sup>. The expression of over 20 different genes in testis would suggest the existence of an equally large number of extracellular signals, and the possible coupling to different effector pathways. It also raises the question of whether a single sperm cell actually expresses the whole set, or only a subset, of the receptor genes.

If these receptors are effectively expressed and functional, it would mean that evolution has selected a common gene family to encode the sensors involved in the chemotaxis of the organism during both the diploid and haploid portions of its life cycle. It would also have obvious implications in the field of male and female infertility and may lead to the development of a new type of contraception. □

Received 23 September; accepted 18 November 1991.

- Libert, F. *et al. Science* **244**, 569–572 (1989).
- Parmentier, M. *et al. Science* **246**, 1620–1622 (1989).
- Buck, L. & Axel, R. *Cell* **65**, 175–187 (1991).
- Libert, F., Vassart, G. & Parmentier, M. *Curr. Opin. Cell Biol.* **3**, 218–223 (1991).
- Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. *A. Rev. Biochem.* **60**, 653–688 (1991).
- Garbers, D. L. *A. Rev. Biochem.* **58**, 719–742 (1989).
- Rait, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 2840–2844 (1991).
- Villanueva Diaz, C., Vadillo Ortega, F. & Kably Ambe, A. *Fert. Steril.* **54**, 1180–1182 (1990).
- Themmen, A. P. N. *et al. Molec. cell. Endocr.* **78**, R7–13 (1991).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1989).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
- Higgins, D. G. & Sharp, P. M. *Gene* **73**, 237–244 (1988).

ACKNOWLEDGEMENTS. We thank J. E. Dumont for his critical interest. This work was supported by the Belgian programme on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's office, Science Policy programming. The scientific responsibility is assumed by the authors. Also supported by grants from the Fonds de la Recherche Scientifique Médicale, and Boehringer Ingelheim. M.P., F.L. and S.S. are Chercheur Qualifié, Chargés de Recherche, and Aspirant of the Fonds national de la Recherche Scientifique of Belgium, respectively. D.E. is an IRSIA fellow.

## Different conformations for the same polypeptide bound to chaperones DnaK and GroEL

Samuel J. Landry, Robert Jordan\*, Roger McMacken\* & Lila M. Gierasch

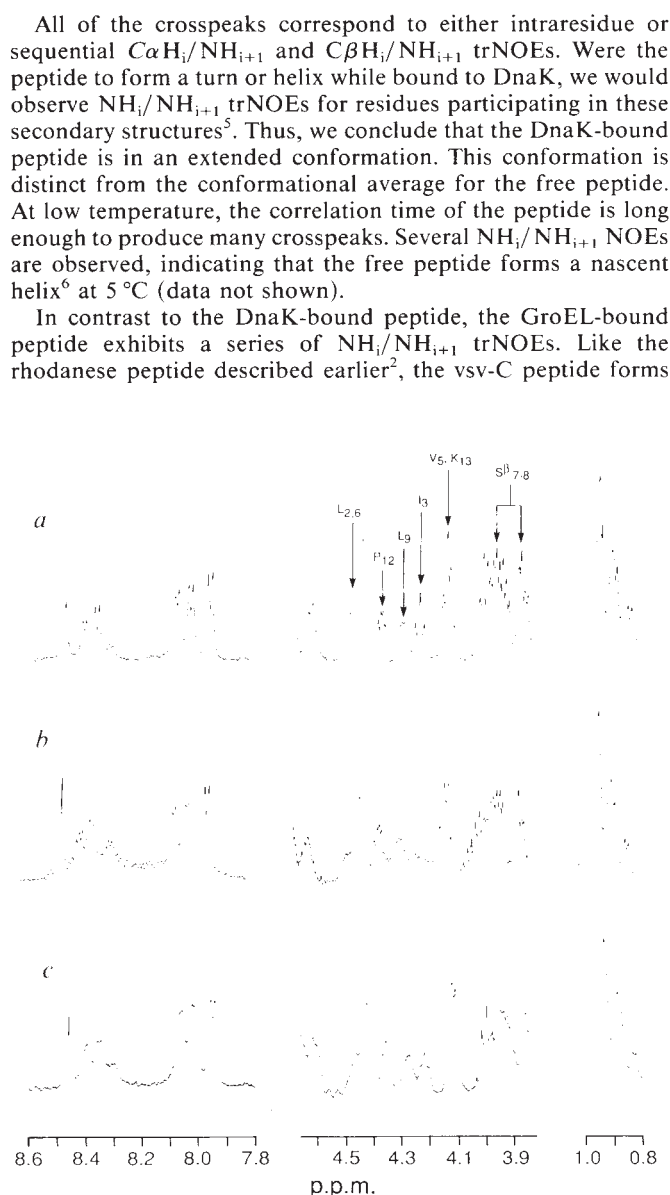
Department of Pharmacology, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75325-9041, USA

\* Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205, USA

**THE proteins DnaK (hsp70) and GroEL (cpn60) from *Escherichia coli* are prototypes of two classes of molecular chaperones conserved throughout evolution<sup>1</sup>. The analysis of transferred nuclear Overhauser effects in two-dimensional NMR spectra is ideally suited to determine chaperone-bound conformations of peptides<sup>2</sup>. The peptide vsv-C (amino-acid sequence KLIGVLSLFRPK) stimulates the ATPase of BiP and Hsc70 (ref. 3) and the intrinsic ATPase of DnaK. The affinity of the vsv-C peptide for DnaK is greatly reduced in the presence of ATP. Here we analyse transferred nuclear Overhauser effects and show that the peptide is in an extended conformation while bound to DnaK but is helical when bound to GroEL. NMR also indicates that the mobility of the peptide backbone is reduced more by binding to DnaK than by binding to GroEL, whereas the side chains are less mobile when bound to GroEL.**

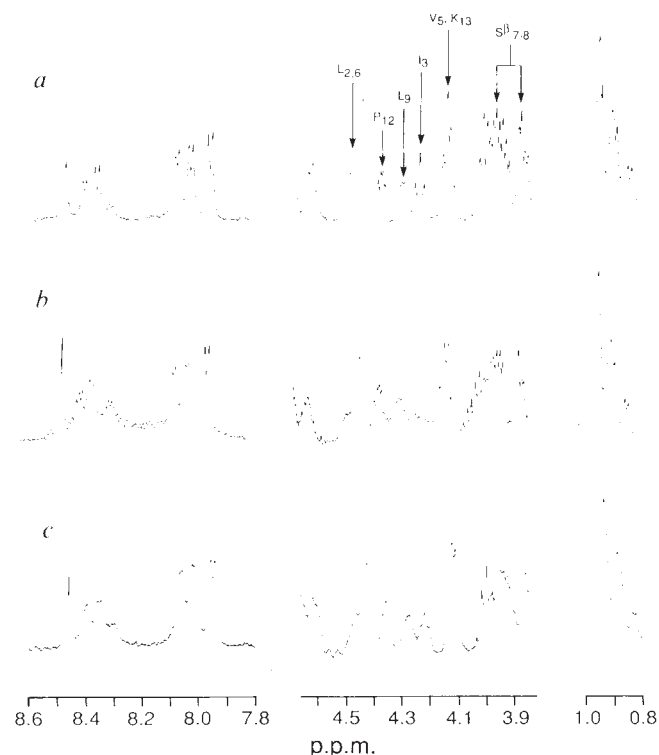
Nuclear Overhauser effects (NOEs) arise from crossrelaxation of magnetization between protons that approach each other closely in space. These 'through space' connectivities are observed as off-diagonal crosspeaks in two-dimensional nuclear Overhauser (2-D NOESY) spectra. In practice, only interproton distances less than about 4 Å produce NOE crosspeaks; thus, the presence or absence of interresidue crosspeaks provides information about the prevailing backbone conformation. The efficiency of crossrelaxation is also dependent on the correlation time of the molecule. In aqueous solution at 25 °C, the vsv-C peptide displays few crosspeaks as a consequence of its rapid tumbling (short correlation time) and conformational flexibility (Fig. 1a). In the presence of the DnaK protein of relative molecular mass 70,000 (*M<sub>r</sub>* 70K), many crosspeaks (transferred NOEs, trNOEs) are observed because the free peptide is in intermediate to fast exchange with the bound peptide<sup>4</sup> (Fig. 1b; see also Fig. 2 legend). While the peptide is bound to DnaK, its crossrelaxation is highly efficient owing to the long correlation time of the large DnaK molecule. After exchange into the free population, relaxation of magnetization is slow. In effect, the free peptide is imprinted with the pattern of magnetization acquired while in the bound conformation.





**METHODS.** The peptide was synthesized with a Milligen 9050 continuous flow peptide synthesizer using 9-fluorenyl-methoxycarbonyl (Fmoc) chemistry and purified by reversed-phase HPLC. DnaK (ref. 12) and GroEL (ref. 2) were purified to homogeneity by modification of previously described procedures. Chaperones were concentrated to about 1.5 mM by centrifugation in a Centricon-30 ultrafiltration device (Amicon) before dilution into samples of the peptide; 3 mM peptide, 100  $\mu$ M DnaK or 25  $\mu$ M GroEL monomer. NMR spectra were recorded on a Varian VXR500 NMR spectrometer operating at 500 MHz at 25 °C. Resonances were assigned using total correlation (TOCSY) and NOESY 2D-spectra which were obtained with standard pulse schemes and processed with the program FTNMR (Hare Research, Seattle) on a Sun4/260 workstation. For NOESY spectra, a total of 256 free induction decays (FIDs) were acquired with sweep width 6,000 Hz, 2,048 points and 32 scans per FID. The mixing time was either 300 ms (*a* and *b*) or 150 ms (*c*). The data were symmetrized<sup>13</sup> for clarity; all of the crosspeaks shown were present in unsymmetrized spectra.

In contrast to the DnaK-bound peptide, the GroEL-bound peptide exhibits a series of  $\text{NH}_i/\text{NH}_{i+1}$  trNOEs. Like the rhodanese peptide described earlier<sup>2</sup>, the vsv-C peptide forms



**METHODS.** Samples were prepared as described in the legend to Fig. 1 except in *b* where the concentration of DnaK was 165  $\mu$ M. NMR methods were as described in the legend to Fig. 1, except that the FIDs were acquired with 128,000 points and 128 scans. Data sets were zero-filled to 260,000 points, and 0.5 Hz line-broadening was applied to all spectra to reduce noise.

an  $\alpha$  helix in association with GroEL (Fig. 1c). Thus, the same peptide adopts distinctly different conformations when it binds to the bacterial hsp70 and cpn60 molecular chaperones. This behaviour may be general as the rhodanese peptide also adopts an extended conformation on binding to DnaK (data not shown).

Further information on the peptide's mode of interaction with DnaK was obtained from differential line-broadening effects. Low mobility of protons in the peptide-protein complex contributes to resonance line broadening and can indicate greater participation in the binding interaction<sup>7</sup>. Peptide:chaperone ratios were selected so that broadening of NH resonances was roughly equal in the presence of DnaK and GroEL (Fig. 2). This normalization compensates for differences in peptide-binding affinity and molecular size of the chaperones and allows assessment of line broadening of non-NH resonances relative to that of NH resonances. Two of the  $\alpha$ H resonance lines, those of Ile3 and Val5, are considerably more broadened in the presence of DnaK than in the presence of GroEL. Leu2, Gly4 and Leu6  $\alpha$ H resonance lines may also be differentially broadened by DnaK, but this cannot be determined unambiguously because of other overlapping resonances. We conclude that the region of Ile3–Val5 has the strongest interaction with DnaK. Furthermore, the interaction with DnaK is mediated in part by hydrogen bonds to the backbone in this region of the peptide because the  $\alpha$ H resonances are broadened more than those of side chains in the same region (data not shown). By contrast, side-chain proton resonances, most notably the Ser $\beta$ s and the methyl groups, are more broadened in the spectrum of peptide with GroEL. Moreover, the interaction with GroEL is less localized. The helical mode of binding presents a hydrophobic surface composed of side chains spanning most of the sequence.

It has been proposed that hsp70 proteins use the energy of ATP hydrolysis to dissociate protein aggregates<sup>8</sup>. Hydrolysis of ATP by BiP and Hsc70 has been thought to be coupled to binding and release of small peptides from these hsp70 family chaperones<sup>3</sup>. TrNOE crosspeaks formed in the presence of DnaK decreased in size or disappeared upon addition of ATP

(Fig. 3). Intensity gradually began to return during prolonged incubation as the level of ATP drops (data not shown). These findings confirm that peptide affinity for DnaK is greatly reduced when ATP is present (R.J. and R.M., unpublished results).

Peptide binding to DnaK is probably mediated by hydrogen bonds between DnaK and the peptide backbone, as well as by hydrophobic contacts and possible salt bridges with the side chains. In general, this mode of binding is similar to that of antigens to major histocompatibility complex (MHC) molecules<sup>9</sup>, and the C-terminal domain of Hsc70 has been modelled after the structure of the MHC peptide-binding domain<sup>10,11</sup>. Peptide binding to GroEL seems to rely more on side-chain hydrophobicity, although the innate capacity of the peptide to form an  $\alpha$  helix is also important<sup>2</sup>. □

Received 16 September; accepted 14 November 1991.

1. Morimoto, R. I., Tissières, A. & Georgopoulos, C. (eds) *Stress Proteins in Biology and Medicine* (Cold Spring Harbor Laboratory, New York, 1990).
2. Landry, S. J. & Gierasch, L. M. *Biochemistry* **30**, 7359–7362 (1991).
3. Flynn, G. C., Chappell, T. G. & Rothman, J. E. *Science* **245**, 385–390 (1989).
4. Clote, G. M. & Gronenborn, A. M. *J. magn. Reson.* **48**, 402–417 (1982).
5. Wüthrich, K., Billeter, M. & Braun, W. *J. molec. Biol.* **180**, 715–740 (1984).
6. Dyson, H. J., Rance, M., Houghton, R. A., Lerner, R. A. & Wright, P. E. *J. molec. Biol.* **201**, 161–200 (1988).
7. Ni, F., Konishi, Y., Frazier, R. B. & Sheraga, H. A. *Biochemistry* **28**, 3082–3094 (1989).
8. Pelham, H. R. B. *Cell* **46**, 959–961 (1986).
9. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **323**, 321–325 (1991).
10. Flajnik, M. F., Camilo, C., Kramer, J. & Kasahara, M. *Immunogenetics* **33**, 295–300 (1991).
11. Rippman, F., Taylor, W. R., Rothbard, J. B. & Green, N. M. *EMBO J.* **10**, 1053–1059 (1991).
12. Mensa-Wilmot, K. et al. *J. biol. Chem.* **264**, 2853–2861 (1989).
13. Baumann, R., Wider, G., Ernst, R. R. & Wüthrich, K. *J. magn. Res.* **44**, 402–406 (1981).

ACKNOWLEDGEMENTS. We thank T. Higashijima and M. Washabaugh for reviewing this manuscript. This work was supported by grants from the NIH (to R.M. and L.M.G.) and from the Robert A. Welch Foundation (to L.M.G.). S.J.L. is a postdoctoral fellow of the American Cancer Society.

## A transcriptional hierarchy involved in mammalian cell-type specification

Calvin J. Kuo, Pamela B. Conley, Lei Chen, Frances M. Sladek\*, James E. Darnell Jr\* & Gerald R. Crabtree

Howard Hughes Medical Institute, Unit in Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305-5425, USA

\* Laboratory of Molecular Cell Biology, The Rockefeller University, New York, New York 10021, USA

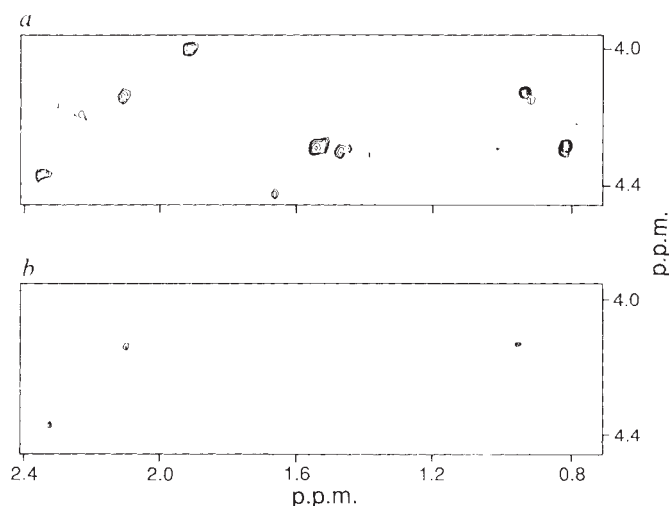


FIG. 3 NOESY spectra showing the  $\alpha$ H/aliphatic region of the peptide with DnaK (a) and with DnaK and ATP (b). Addition of ATP reduces the affinity of peptide for DnaK and trNOE crosspeaks are consequently lost or reduced. METHODS. The sample was the same as in Fig. 2b (3 mM peptide, 165  $\mu$ M DnaK) NMR methods were as in Fig. 1. In b,  $\text{MgCl}_2$  and Na-ATP (pH 7.0) were added to final concentrations of 12.5 mM and 10 mM, respectively. Phosphate anion partially inhibits the DnaK ATPase at the concentration used in this experiment (R.J. and R.M., unpublished results). To observe reduced peptide binding, a high concentration of ATP may be needed to compensate for inhibition by phosphate. Sodium azide (0.5%) was included in the sample after addition of ATP to inhibit microbial growth.

ALTHOUGH transcriptional hierarchies have been extensively studied in invertebrates, their involvement in mammalian cell-type specification is poorly understood<sup>1,2</sup>. Here we report a hepatocyte transcriptional cascade suggested by the expression patterns of hepatic transcription factors in dedifferentiated hepatomas and hepatocyte: fibroblast hybrids in which the liver phenotype was extinguished<sup>3–6</sup>. These results indicated that the homeoprotein hepatocyte nuclear factor-1 $\alpha$  (HNF-1 $\alpha$ )<sup>7–8</sup>, and HNF-4, a member of the steroid hormone receptor superfamily<sup>9</sup>, were regulated coordinately or in a hierarchy by a higher-order locus, independently of other hepatic transactivators. HNF-4 was implicated as an essential positive regulator of HNF-1 $\alpha$ , as deletion of an HNF-4 binding site in the HNF-1 $\alpha$  promoter abolished promoter activity, and HNF-4 potentially transactivated the HNF-1 $\alpha$  promoter in cotransfection assays. Moreover, genetic complementation of dedifferentiated hepatomas with HNF-4 complementary DNA rescued expression of endogenous HNF-1 $\alpha$  messenger RNA and DNA-binding activity. Our studies therefore define an HNF-4  $\rightarrow$  HNF-1 $\alpha$  (4  $\rightarrow$  1 $\alpha$ ) transcriptional hierarchy operative in differentiated hepatocytes but selectively inhibited by an extinguishing locus and somatic mutations which antagonize the liver phenotype.



Variant, dedifferentiated hepatomas with somatic mutations or fixed epigenetic alterations were used to correlate the expression of hepatic transcription factors with the loss of the liver phenotype. From the differentiated rat hepatoma Fao, a morphological variant was derived, C2 (ref. 3), which lacks expression of several hepatic mRNAs<sup>10,11</sup> and the hepatic HNF-1 $\alpha$  and HNF-4 transactivators (Fig. 1). A rare ( $10^{-9}$ ) C2 revertant, termed Rev7, selected by growth in glucose-free medium, has restored hepatocyte morphology<sup>4</sup>, liver-specific and gluconeogenic gene expression<sup>4,10</sup>, as well as HNF-1 $\alpha$  and HNF-4 (Fig. 1) mRNA and DNA-binding activity. We have isolated a morphological Rev7 variant, designated R8, which again lacks liver-specific gene expression, HNF-1 $\alpha$  and HNF-4 (Fig. 1, and data not shown). The hepatic C/EBP and HNF-3 transactivators did not exhibit such variation in this cell panel, in agreement with previous results<sup>12,13</sup> (Fig. 1a). As expected, HNF-1 $\alpha$ , HNF-3 and HNF-4 were expressed at high levels in rat liver and the Hepa 1a and HepG2 differentiated hepatomas (Fig. 1, and data not shown). The stability and frequency of these dedifferentiated cells ( $10^{-6}$ ) and revertants ( $10^{-9}$ ) are consistent with mutagenesis of a genetic locus<sup>3,4</sup> coordinately regulating HNF-1 $\alpha$  and HNF-4.

Genetic loci regulating the hepatocyte phenotype have also been defined by correlating retained chromosomes in rat Fao  $\times$  mouse fibroblast somatic cell hybrids (FF) with extinction of liver gene expression. The FF5-1 and FF3-3 hybrid lines have retained mouse chromosome 1 and associated extinguishing loci, and have consequently repressed expression of numerous hepatic HNF-1 $\alpha$ -dependent and -independent genes<sup>5,6</sup>. Both FF5-1 and FF3-3 hybrids exhibited coordinate loss of HNF-1 $\alpha$  and HNF-4, whereas C/EBP and HNF-3 expression were reminiscent of the parental Fao line (Fig. 1). Thus, two independent approaches, variant hepatomas and extinguished hybrids, suggested the existence of genetic loci regulating HNF-1 $\alpha$  and HNF-4 independent of other hepatic transcription factors.

To define regulatory interactions between HNF-1 $\alpha$ , HNF-4 and extinguishing loci, we characterized the murine HNF-1 $\alpha$  (mHNF-1 $\alpha$ ) promoter. HNF-1 $\alpha$  transcription start sites clustered about the 5' end of previously isolated<sup>14</sup> mHNF-1 $\alpha$  cDNA (designated '+1') (Fig. 2a, b, and data not shown). Correctly initiated transcripts were detected in liver and kidney, but not brain or heart, paralleling the tissue distribution of HNF-1 $\alpha$  (refs 14–16) (Fig. 2b). Sequential 5' deletions indicated that a construct 110UCAT, containing bases -110 to +140, was sufficient to direct tissue-specific expression of a linked chlor-

amphenicol acetyltransferase gene (data not shown). Hepa 1a hepatoma nuclear proteins produced four DNase I footprints, designated A, B1, B2 and C, on this minimal fragment (Fig. 2a, c). DNase I protection between the B1 and A footprints (Fig. 2c) was not consistently observed on either strand.

The C-site factor was identified as HNF-4 by efficient association of HNF-4 antibodies<sup>9</sup> with C-site protein-DNA complexes to produce a ternary 'supershift' complex of reduced mobility (Fig. 2d, and data not shown). Additionally, the B2- and B1-site proteins were identified, respectively, as HNF-3 $\alpha\beta\gamma$ , mammalian homologues of the *Drosophila* homeotic protein from *fork head*<sup>17,18</sup> and AP-1, a protein kinase C-responsive transcription factor. Indeed, *in vitro* translated HNF-3 $\alpha$ , -3 $\beta$  and -3 $\gamma$  associated equally well with both the B2-site and a transthyretin (TTR) promoter HNF-3 site<sup>17</sup>, and *in vitro* translated *c-fos* efficiently heterodimerized with *in vitro* translated *c-jun*, *junB* or *junD* on the B1-site, characteristic of AP-1 (ref. 19; Fig. 2f). These assignments have been confirmed by gel-shift comigration and cross-competition experiments (data not shown). We have been unable to identify the A-site factor, which is present in hepatocyte lines but not liver tissue.

In functional studies, HNF-4 behaved as an essential positive regulator of the HNF-1 $\alpha$  promoter. Deletion of the C/HNF-4 site abolished (93–95%) HNF-1 $\alpha$  promoter activity in transfected HepG2 cells (Fig. 3a), revealing an absolute HNF-4 dependence, and suggesting that the absence of HNF-4 in dedifferentiated and extinguished hepatomas was causally related to the absence of HNF-1 $\alpha$ . By contrast, B2/HNF-3 site deletion only produced a 32% reduction, consistent with the poor correlation between HNF-3 and HNF-1 $\alpha$  expression in the cell panel (Figs 3a and 1a). Deletion of the A-site or a control site 'AR' between the HNF-3 and HNF-4 sites was associated with increased activity (Fig. 3a, and data not shown).

Moreover, HNF-4 transactivated the HNF-1 $\alpha$  promoter 3.8-fold in transfected Jurkat T-lymphocytes but 14.8-fold in R8, indicating that the selective absence of HNF-4 in R8 cells was a limiting factor for HNF-1 $\alpha$  expression (Fig. 3b). By contrast, HNF-3 was relatively ineffective in either Jurkat or R8 cells (Fig. 3b), corroborating results from deletion mutants (Fig. 3a) and cell lines (Fig. 1) suggesting that HNF-4 was the dominant regulator of the HNF-1 $\alpha$  promoter. In Jurkat cells, the phorbol ester analogue PdBu induced AP-1 activity (data not shown, and ref. 18) and enhanced HNF-4 transactivation 3.5-fold to 13.2-fold (Fig. 3b). PdBu treatment alone was not stimulatory (data not shown), implying that AP-1 is permissive but not

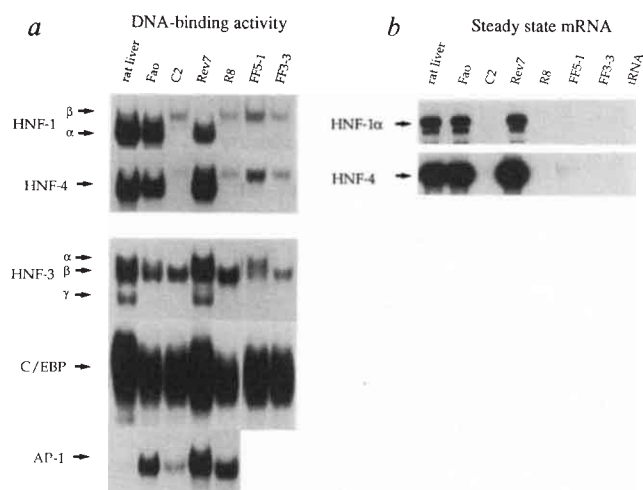


FIG. 1 a, Selective absence of HNF-1 $\alpha$  and HNF-4 gel mobility shift complexes in variant hepatocyte cell lines and extinguished hepatocyte: fibroblast hybrids. The positions of HNF-1 $\alpha$  versus HNF-1 $\beta$ , and HNF-3 $\alpha\beta\gamma$  are indicated by arrows. b, RNase protection analysis of HNF-1 $\alpha$  and HNF-4 mRNA in variant hepatomas and extinguished hepatocyte: fibroblast hybrids. Arrows indicate correctly protected fragments of 240 nucleotides (HNF-1 $\alpha$ ) and 179 nucleotides (HNF-4).

METHODS. Nuclear extract (5  $\mu$ g) from the indicated cell lines or rat liver were incubated with <sup>32</sup>P-end-labelled double-stranded binding site oligomers under standard gel mobility shift assay conditions<sup>10,25</sup>. To distinguish between HNF-1 $\alpha$  and HNF-1 $\beta$ , HNF-1 $\beta$ -specific antiserum was included in the binding reaction, generating an HNF-1 $\beta$  'supershift' complex of decreased mobility<sup>25,26</sup>. HNF-1 $\alpha\beta$ , C/EBP and AP-1 oligos have been previously described<sup>10,27,28</sup>. Sequences of HNF-4 and HNF-3 oligos are HNF-4 sense 5'-TCGAGGCTGAAGTCCAAAGTTTCAGTCCCTTCG-3', HNF-4 antisense 5'-TCGACGAAGGGACTGAACTTTGGACTTCAGCC-3', HNF-3 sense 5'-TCGAGCTCCAATGTAAACAGAGCAGG-3', HNF-3 antisense 5'-TCGACCTGCTCTGTTTACAT-TGGAGC-3'. Total RNA was isolated by the guanidine isothiocyanate method<sup>24</sup> and 15  $\mu$ g samples were hybridized to HNF-1 $\alpha$  and HNF-4 antisense riboprobes<sup>25</sup>. The HNF-1 $\alpha$  riboprobe was generated by T3 transcription of *Pst*I-linearized p1.1 (ref. 14). To generate the HNF-4 riboprobe, a 179-nucleotide *Pvu*II-*Nhe*I (+613 to +802) fragment of rat HNF-4 cDNA<sup>9</sup> was subcloned into *Sma*I-*Xba*I-digested plasmid pBSKS(-), religated with *Eco*RI and transcribed with T7.

required for HNF-1 $\alpha$  expression, as is the cyclic AMP-responsive factor CREB for the anterior pituitary homeoprotein GHF-1/Pit-1 (refs 20, 21). The action of AP-1 on the HNF-4-activated HNF-1 $\alpha$  promoter suggests that HNF-1 $\alpha$  may be developmentally regulated by environmental/inductive signals.

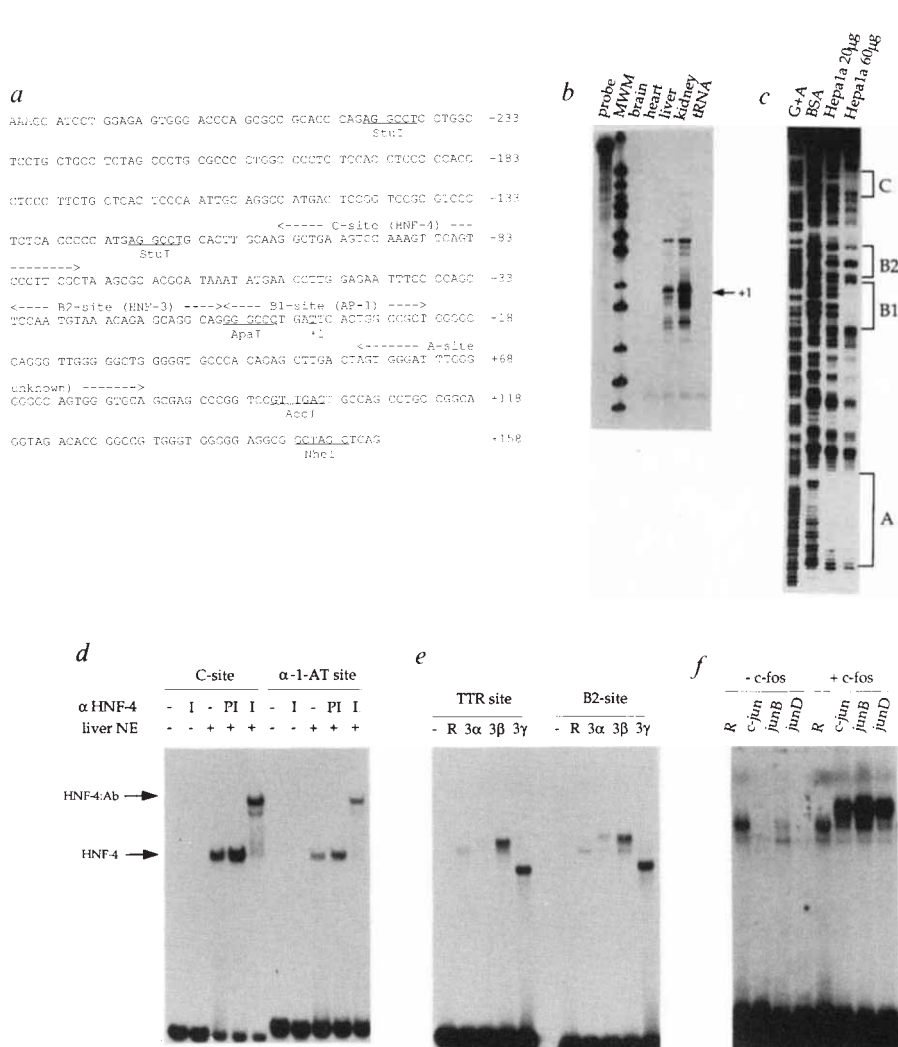
An HNF-4  $\rightarrow$  HNF-1 $\alpha$  genetic hierarchy was established by genetic complementation of the HNF-4(-) HNF-1 $\alpha$ (-) R8 hepatoma with HNF-4. Several HNF-4-expressing neomycin-resistant lines were obtained, with the highest-expressing clone propagated for further study (R8::HNF4). HNF-4 expression was not observed in the parental R8 line, or in six control clones transfected with neomycin-resistance marker alone (R8::neo), although abundant immunoreactive HNF-4 DNA-binding activity was present in R8::HNF4 (Figs 4a and 1, and data not shown). Notably, HNF-4-expressing cells activated the previously silent, endogenous HNF-1 $\alpha$  gene, as evidenced by rescue of HNF-1 $\alpha$  mRNA and supershifted DNA-binding complexes in R8::HNF-4 (Fig. 4b, c) and C2::HNF-4 (C.J.K. and G.R.C.,

unpublished results). No HNF-1 $\alpha$  mRNA/supershift complexes were detected in the parental R8 line or control neo clones, which contained only endogenous HNF-1 $\beta$  (Figs 4b, c and 1). Arguing against, but not excluding, reciprocal regulation of HNF-4 by HNF-1 $\alpha$ , lines overexpressing HNF-1 $\alpha$  did not induce HNF-4 (Fig. 4a, and data not shown), and preliminary results have not revealed HNF-1 $\alpha$  binding sites in the HNF-4 promoter (W. Zhong, F.M.S. and J.E.D.Jr, unpublished results).

We have described several regulatory interactions between hepatic transcription factors (Fig. 4d). In contrast to other cell-type-specific transcription factors (compare Pit-1/GHF-1 and GATA-1; refs 20-22), we have not obtained evidence for HNF-1 $\alpha$  promoter autoregulation. Rather, the liver-enriched factors HNF-3 and HNF-4 confer promoter tissue-specificity, with AP-1 acting permissively to integrate cell-extrinsic signals. Our inspection of dedifferentiated and extinguished hepatomas further suggests that higher-order loci regulate HNF-1 $\alpha$  and HNF-4 independently of C/EBP and HNF-3. Genetic evidence

FIG. 2 Characterization of the murine HNF-1 $\alpha$  promoter and associated DNA-binding proteins. *a*, Nucleotide sequence of the murine HNF-1 $\alpha$  promoter and 5'UTR. The predominant transcription start site, corresponding to the 5' end of previously isolated mHNF-1 $\alpha$  cDNA<sup>14</sup>, is underlined and designated '+1'. Binding sites for nuclear factors and restriction sites used in the present studies are underlined. *b*, RNase protection mapping of 5' ends of mHNF-1 $\alpha$  transcripts in different tissues. The major 5' end is designated '+1'. 'Probe', undigested riboprobe; 'MWM', molecular weight markers (*MspI* digest of pBR322). Similar results have been obtained by primer extension. *c*, DNase I footprinting of the HNF-1 $\alpha$  promoter and 5'UTR with nuclear extracts from the Hepa 1a differentiated hepatoma or BSA (20  $\mu$ g). The B2-site is characterized by generation of a hypersensitive site at -22; protection from DNase I was observed for the C, B1 and A footprints. 'G + A', Maxam and Gilbert piperidine cleavage of the footprinting probe. *d*, Reactivity of HNF-4 antiserum with C-site and  $\alpha$ -1-AT HNF-4 site gel shift complexes. Specific formation of antibody HNF-4/DNA complexes was observed with immune sera (I) and not preimmune sera (PI). *e*, *In vitro* translated HNF-3 $\alpha$ , HNF-3 $\beta$  and HNF-3 $\gamma$  form identical gel shift complexes with either transthyretin (TTR) HNF-3 site or B2-sites. A nonspecific complex (asterisk) was observed in unprogrammed reticulocyte extract (R). *f*, *In vitro* translated *c-fos* and *jun* family members heterodimerize on the B1 site in gel-shift assays. Nonspecific complexes from unprogrammed reticulocyte extract (R) are indicated (\*). Specific complex formation depended on mixing of translated *fos* and *jun* ('+c-fos' lanes), characteristic of AP-1.

**METHODS.** A mouse liver  $\lambda$ J1 genomic library was screened under high stringency conditions<sup>14</sup> with <sup>32</sup>P-labelled *EcoRI*-*XhoI* fragment of mHNF-1 cDNA (+1 to +220)<sup>14</sup>, with isolation of phage with ~6.5 kb of 5' genomic sequences. Sequencing of ~400 bp was done by the Sanger dideoxy termination method<sup>14</sup>. To identify HNF-1 $\alpha$  transcription start sites, a *StuI*-*AccI* genomic HNF-1 $\alpha$  fragment (-118 to +103) (Fig. 2a) was subcloned into *SmaI*-*AccI*-linearized pBS KS(-) (pBSrbr). T3 transcription of the *XbaI*-linearized construct produced an antisense riboprobe spanning the putative transcription start site, which was hybridized at 56 °C against 15  $\mu$ g total tissue RNA or transfer RNA, in RNase protection assay<sup>25</sup>. DNase I footprinting and piperidine (G + A) cleavage were as previously described<sup>10</sup>. The footprinting probe was generated from pBSrbr by digestion with *Asp*718 and *SacII* and labelling with <sup>32</sup>P- $\alpha$ -TTP/Klenow. HNF-4 immune and preimmune sera, and human  $\alpha$ -1-AT



HNF-4, rat TTR HNF-3 and human interleukin-2 AP-1 binding site probes have been described elsewhere<sup>9,17,28</sup>; c, B2 and B1 double-stranded sites were synthesized as in *a*. *In vitro* transcription/translation was done under standard conditions<sup>26</sup>, with 0.5–2  $\mu$ l of a 50- $\mu$ l translation added into the gel shift binding reaction. HNF-3 transcripts were prepared as described<sup>17</sup>, and *fos* and *jun* transcripts as described elsewhere (J. Northrop and G.R.C., manuscript in preparation).



FIG. 3 Functional analysis of the HNF-1 $\alpha$  promoter by *a*, transfection of binding site deletion mutants into the differentiated hepatoma HepG2 and *b*, transactivation with combinations of CMVHNF4, CMVHNF3 and PdBu treatment in Jurkat cells (top) and R8 cells (bottom).

**METHODS.** Cells were transfected by electroporation (Jurkat: 0.22 kV, 960  $\mu$ F, Bio-Rad Gene Pulser) or calcium phosphate treatment (HepG2, R8: 18 h), and collected 48 h after initial treatment. PdBu (50 ng ml $^{-1}$ ) was added as appropriate 8 h before collecting. No PdBu effect was observed in R8 cells, probably because of high endogenous levels of AP-1, whereas >10-fold AP-1 induction in Jurkat cells under these conditions has been confirmed (Fig. 1*a* and data not shown). Deletional studies used 15  $\mu$ g CAT reporter and 1.5  $\mu$ g RSVluc, and transactivation studies used 7.5  $\mu$ g of each expression vector, 7.5  $\mu$ g CAT reporter, 1.5  $\mu$ g RSVluc and pBSKS(-) as appropriate. CAT and luciferase assays have been previously described<sup>26</sup>, and CAT activity was normalized to luciferase values. Points represent the results of independent experiments with the average of three experiments indicated by bars. The reporter gene 110UCAT was constructed by ligating an HNF-1 $\alpha$  genomic fragment (*StuI*-*NheI*, -118 to +140) into the promoter-less KSCAT plasmid (C.J.K. and P.B.C., unpublished results). Binding site deletions in 110UCAT were constructed by PCR overlap extension<sup>29</sup> and removed nucleotides -98 to -79 ( $\Delta$ C), -30 to -11 ( $\Delta$ B2) and -67 to -46 ( $\Delta$ AR). CMVHNF4 was derived from the pLEN4S HNF-4 expression vector<sup>9</sup> by release of HNF-4 cDNA with *Bam*HI, addition of *Eco*RI linkers, and ligation into the CMV expression vector pBC5. CMVHNF3 $\gamma$  was obtained from E. Lai; similar results have been obtained with HNF-3 $\alpha$  and HNF-3 $\beta$  expression vectors (data not shown).

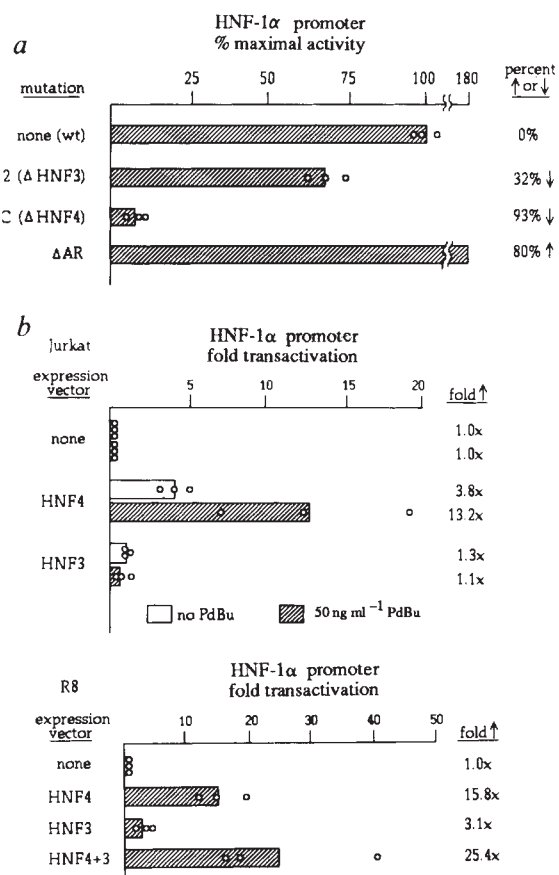
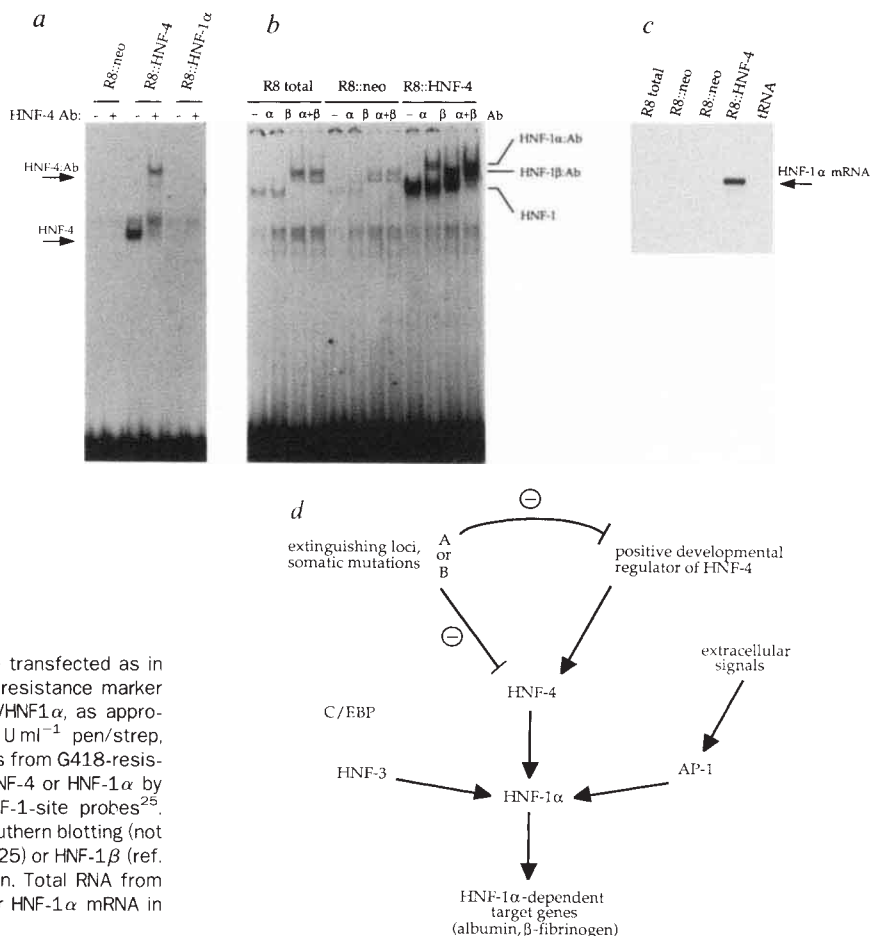


FIG. 4. Rescue of endogenous HNF-1 $\alpha$  expression by HNF-4 cDNA. *a*, Overexpression of immunoreactive HNF-4 in R8 cells. The position of HNF-4:DNA and antibody:HNF-4:DNA complexes are indicated. R8:neo, R8:HNF-4 and R8:HNF-1 $\alpha$  refer to R8 clones stably transfected with neomycin-resistance marker, CMVHNF4 + neo and CMVHNF1 $\alpha$  + neo, respectively. *b*, HNF-4 rescues endogenous HNF-1 $\alpha$  DNA-binding activity in R8 cells. HNF-1 (total) and Ab:HNF-1 $\alpha$  or Ab:HNF-1 $\beta$  complexes are indicated by arrows. HNF-1 $\alpha$  supershift complexes are only apparent in R8:HNF-4. Note total supershift of parental R8 and R8:neo HNF-1 binding activity with anti-HNF-1 $\beta$  serum, but complexes resistant to HNF-1 $\beta$  serum in R8:HNF-4. These 1 $\beta$ -resistant complexes are subsequently supershifted by HNF-1 $\alpha$  antiserum, again demonstrating HNF-1 $\alpha$  expression. *c*, Rescue of HNF-1 $\alpha$  mRNA expression in R8:HNF-4. Specific RNase protection of HNF-1 $\alpha$  antisense riboprobe by R8:HNF-4 RNA is indicated by the arrow. *d*, Summary of deduced regulatory interactions.

**METHODS.** Dishes (150-mm diameter) of R8 cells were transfected as in Fig. 3 with 4  $\mu$ g *XmnI*-linearized CX12neo.tr neomycin resistance marker with or without 40  $\mu$ g *Sall*-linearized CMVHNF4 or CMVHNF1 $\alpha$ , as appropriate. Drug selection was in DMEM, 10% FCS, 100 U ml $^{-1}$  pen/strep, 560  $\mu$ g ml $^{-1}$  G418 (active drug) (GIBCO). Nuclear extracts from G418-resistant clones were prepared<sup>25</sup> and 7.5  $\mu$ g analysed for HNF-4 or HNF-1 $\alpha$  by gel mobility shift assay with C-site (Fig. 2*a*) and HNF-1-site probes<sup>25</sup>. Integration of appropriate plasmids was confirmed by Southern blotting (not shown). Where required, 1  $\mu$ l HNF-4 (ref. 9), HNF-1 $\alpha$  (ref. 25) or HNF-1 $\beta$  (ref. 26) specific antisera were added to the binding reaction. Total RNA from transfected clones was isolated and 20  $\mu$ g analysed for HNF-1 $\alpha$  mRNA in RNase protection assay (see Fig. 1*a*).



for such a regulatory demarcation has been observed in homozygous *lethal albino* mice, which have severely impaired hepatic gene expression, and in which HNF-1 $\alpha$  and HNF-4 but not other hepatic factors are selectively repressed<sup>23</sup>. The absolute HNF-4-dependence of the HNF-1 $\alpha$  promoter, combined with the ability of HNF-4 to rescue endogenous HNF-1 $\alpha$  expression, indicates that HNF-4 is an essential regulator of HNF-1 $\alpha$  and that the absence of HNF-4 and HNF-1 $\alpha$  are causally related in mouse chromosome 1-containing extinguished hybrids and mutant hepatomas. We therefore suggest that extinguishing loci and somatic mutations indirectly reduce HNF-1 $\alpha$  expression

through HNF-4. The locus responsible for HNF-4/HNF-1 $\alpha$  extinction is unlikely to be the chromosome 1 extinguishing locus *tse-2*, because *tse-2*-expressing minimal hybrids have been reported to express HNF-1 $\alpha$  (ref. 24). Thus, an uncharacterized extinguishing locus probably represses HNF-4 either directly or through a positive developmental regulator of HNF-4 (Fig. 4d). The definition of mechanisms underlying the absence of HNF-4 mRNA in variant and hybrid hepatomas (Fig. 1b) should grant insight into mechanisms of phenotypic extinction, as well as the interplay between positive and negative developmental strategies. □

Received 23 September; accepted 5 November 1991.

- Winslow, G. M., Hayashi, S., Krasnow, M., Hogness, D. S. & Scott, M. P. *Cell* **57**, 1017–1030 (1989).
- Hayashi, H. & Scott, M. P. *Cell* **63**, 883–894 (1990).
- Deschatrette, J. & Weiss, M. C. *Biochimie* **56**, 1603–1611 (1974).
- Deschatrette, J., Moore, E. E., Dubois, M. & Weiss, M. C. *Cell* **19**, 1043–1051 (1980).
- Killary, A. M. & Fournier, R. E. K. *Cell* **38**, 523–534 (1984).
- Chin, S. C. & Fournier, R. E. K. *Molec. cell. Biol.* **9**, 3736–3743 (1989).
- Frain, M. *et al.* *Cell* **59**, 1179–1187 (1989).
- Baumhueter, S. *et al.* *Genes Dev.* **4**, 372–379 (1990).
- Sladek, F. M., Zhong, W., Lai, E. & Darnell, J. E. *Genes Dev.* **4**, 2353–2365 (1991).
- Baumhueter, S., Courtois, G. & Crabtree, G. R. *EMBO J.* **7**, 2485–2493 (1988).
- Cereghini, S., Blumenfeld, M. & Yaniv, M. *Genes Dev.* **8**, 957–974 (1988).
- Cereghini, S., Yaniv, M. & Cortese, R. *EMBO J.* **9**, 2257–2263 (1990).
- Herbst, R. S. *et al.* *New Biol.* **3**, 289–296 (1991).
- Kuo, C. J., Conley, P. B., Hsieh, C., Franke, U. & Crabtree, G. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9838–9842 (1990).
- Courtois, G., Morgan, J. G., Campbell, L. A., Fourel, G. & Crabtree, G. R. *Science* **238**, 688–692 (1987).
- De Simone, V. *et al.* *EMBO J.* **10**, 1435–1443 (1991).
- Lai, E., Prezioso, V. R., Tao, W., Chen, W. S. & Darnell, J. E. *Genes Dev.* **5**, 416–427 (1991).
- Lai, E. *et al.* *Genes Dev.* **4**, 1427–1436 (1990).

- Chiu, R. *et al.* *Cell* **54**, 541–552 (1988).
- McCormick, A., Brady, H., Theill, L. E. & Karin, M. *Nature* **345**, 829–832 (1990).
- Chen, R. *et al.* *Nature* **346**, 583–586 (1990).
- Tsai, S.-F., Strauss, E. & Orkin, S. H. *Genes Dev.* **5**, 919–931 (1991).
- Tonjes, R. R., Xanthopoulos, K. G., Darnell, J. E. Jr & Paul, D. *EMBO J.* (in the press).
- Gourdeau, H. & Fournier, R. E. K. *A. Rev. Cell Biol.* **6**, 69–94 (1990).
- Kuo, C. J., Mendel, D. B., Hansen, L. P. & Crabtree, G. R. *EMBO J.* **10**, 2231–2236 (1991).
- Mendel, D. B., Hansen, L. P., Graves, M. K., Conley, P. B. & Crabtree, G. R. *Genes Dev.* **5**, 1042–1056 (1991).
- Lichtsteiner, S., Wuarin, J. & Schibler, U. *Cell* **51**, 963–973 (1987).
- Emmel, E. A. *et al.* *Science* **246**, 1617–1620 (1989).
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. *Gene* **77**, 51–59 (1989).

ACKNOWLEDGEMENTS. We thank M. Flanagan, J. Jorgensen, J. Noordermeer, B. Blakely, H. Blau and M. Krasnow for discussions, E. Lai for HNF-3 reagents, Y.-H. Chien for mouse genomic libraries, C. Sagerstrom for expression vector pBC5 and J. Riegel for CX12neo.tr. This work was supported by the NIH (P.B.C., G.R.C., J.E.D. Jr) and the Howard Hughes Medical Institute (G.R.C.). F.M.S. is an American Cancer Society postdoctoral fellow, and C.J.K. was supported by the Medical Scientist Training Programme at Stanford University School of Medicine.

## Characterization of cDNA for the large subunit of the transcription initiation factor TFIIF

Teijiro Aso, Haren A. Vasavada, Tatsuo Kawaguchi\*, F. Joseph Germino†, Subinay Ganguly‡, Shigetaka Kitajima\*, Sherman M. Weissman§ & Yukio Yasukochi\*

Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

† Department of Medicine and ‡ Department of Molecular Genetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan

At least six chromatographically resolvable general transcription factors may participate in accurate initiation by RNA polymerase II in HeLa cell-derived systems<sup>1–3</sup>. TFIIF (also termed FC, RAP30/74 and  $\beta/\gamma$ ) can bind directly to RNA polymerase II in solution<sup>4–6</sup> and decrease the affinity of RNA polymerase II for nonspecific DNA<sup>7</sup>. From studies on the kinetics of transcription initiation<sup>1</sup>, on the composition of transcription initiation complexes fractionated by acrylamide gel electrophoresis<sup>8</sup>, and on template competition experiments<sup>9</sup>, TFIIF is known to act at an intermediate stage in initiation complex formation. It acts after TFIID firmly associates with DNA, but coincidentally with or immediately after RNA polymerase II binding to DNA, and before the recruitment of factor TFIIE. TFIIF may<sup>6</sup> or may not<sup>4,5</sup> have DNA helicase activity. The small subunit (RAP30) of TFIIF has been cloned<sup>6</sup> and shows some amino-acid sequence homology to bacterial  $\sigma$  factors. We have partially sequenced the RAP74 protein from purified HeLa cells, cloned its complementary DNA and shown that its translation product can interact with RAP30 *in vitro* as

well as *in vivo*. The cDNA predicts an amino-acid sequence that lacks obvious DNA or RNA helicase motifs. It has regions rich in charged amino acids, including segments containing a higher content of acidic amino acids than are found in strong transcriptional activators such as VP16 (ref. 10).

A full-length cDNA clone of RAP74 containing a 2.3-kilobase (kb) insert detected a single RNA band of 2.8 kb on the northern blot (Fig. 1a). This clone and two other shorter clones were sequenced. The overlapping segments from all three clones had identical sequences. The predicted open reading frame encoded a protein of 517 amino acids with a relative molecular mass of 58,000. This protein contained all four tryptic peptides obtained from RAP74 purified from HeLa cell extracts (Fig. 1b). The cDNA clone, when expressed in bacteria using the pET3d vector<sup>11</sup>, had an electrophoretic mobility close to that observed for RAP74 from HeLa cells (data not shown). The amino-acid sequence contained no conventional adenosine triphosphate binding motif<sup>12</sup>. Consistent with reports that RAP74 is phosphorylated *in vivo*, the amino-acid sequence contains multiple potential phosphorylation sites for both protein kinase C and cyclic AMP-dependent protein kinases, concentrated in the region between residues 350 and 462. Several stretches of acidic amino acids (containing up to 50% aspartic and glutamic acids) are present, particularly in the region from amino acids 184 to 356 (Fig. 1c). These regions are also near stretches of basic residues, particularly lysines, so that the overall pK predicted for the protein would be near neutrality in the absence of phosphorylation. The occurrence of a number of closely spaced clusters of charged residues in the RAP74 protein together with potentially phosphorylated serine residues indicates that electrostatic interactions may be important for RAP74 interactions *in vivo*.

We next produced RAP74 and RAP30 in bacterial cells to test for the ability of these proteins to replace the activity of TFIIF (containing the RAP30/74 complex) in *in vitro* transcription assays. Bacterially produced RAP74 in combination with bacterially produced RAP30 could replace TFIIF in our *in vitro* transcription assay. Bacterial extracts lacking RAP30, RAP74, or both, failed to stimulate transcription (Fig. 2). These results

† Present address: Department of Gene Expression Sciences, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19333, USA.

§ To whom correspondence should be addressed.



FIG. 1 Sequence of human RAP74 cDNA and predicted RAP74 protein. *a*, Northern blot of 2  $\mu$ g poly (A)<sup>+</sup> RNA purified from JY cells (lane 1) and 10  $\mu$ g poly(A)<sup>+</sup> RNA from HeLa cells (lane 2). The positions of the size markers are indicated in kilobases. *b*, Sequence of the cDNA insert and the deduced amino-acid sequence of RAP74. The amino-acid sequences underlined (P1–P4) match sequences that were determined for four tryptic fragments derived from human RAP74. *c*, Schematic representation of the RAP74 cDNA sequence (upper numbers) and deduced amino-acid sequence (lower numbers). The central portion of RAP74 (residues 184–356) is highly acidic (hatched box). The portion containing multiple potential phosphorylation sites (residues 350–462) is shown as a black box. The three thin lines below represent the sequence of the cDNAs JY1, JY3 and JY9.

**METHODS.** RNA samples were fractionated on a 1.0% agarose gel containing 2 M formaldehyde, transferred to a nylon membrane (Hybond, Amersham), baked, and probed with random primer-labelled cDNA containing the entire sequence of RAP74. Blots were washed twice for 15 min at room temperature in 2  $\times$  SSC, 0.1% SDS, and twice for 30 min at 65 °C in 0.2  $\times$  SSC, 0.1% SDS before autoradiography. RAP74, partially purified by consecutive DEAE, P-11 and Sephacryl S-300 column chromatographies<sup>1,5</sup>, was subjected to SDS-PAGE. After electroblotting onto a polyvinylidene difluoride membrane, the RAP74 band was cut out and eluted. The peptides were digested with trypsin and applied onto a Vydac C18 reverse-phase HPLC column (2.1 mm  $\times$  250 mm). The N-terminal sequences of four of the resolved tryptic fragments were determined on an Applied Biosystems sequencer equipped with an on-line HPLC. An oligonucleotide was synthesized corresponding to the most probable estimate<sup>15</sup> of the codons for the amino-acid sequence of peptide P3. A <sup>32</sup>P end-labelled oligonucleotide probe was used to screen a charon BS cDNA library derived from the human lymphoblastoid cell line JY (ref. 16). Three overlapping cDNA clones, JY 1, 3 and 9, with inserts of up to 2.4 kb were isolated. Restriction fragments of these clones were subcloned into pUC119 and Sequenase on both strands using the Sequenase system (USB). The cDNA contained a 1,551-base-pair open reading frame starting with the first methionine. The open reading frame encoded a protein of 517 amino acids.

show that this cDNA does encode the second component of TFIIIF.

To investigate the association of RAP30 and RAP74 *in vivo*, we prepared clones of RAP30 and RAP74 which were fused at their amino termini to the coding sequences for either the GAL4 DNA-binding protein<sup>13</sup> or the transcription activation domain of VP16 (ref. 14) (Fig. 3a). These clones were transfected separately or together into CV1 cells along with a reporter plasmid (G5EC) containing the chloramphenicol acetyltransferase (CAT) gene. The CAT gene was located downstream of a minimal promoter which, in turn, was immediately downstream of a sequence containing five GAL4-binding sites. Interestingly, although RAP74 has two stretches of roughly 40 amino acids, each with a higher density of acidic residues than the acidic activation domain of VP16, GAL4-RAP74 failed to show any transcriptional activation *in vivo*, as measured by its inability to activate transcription of the CAT gene. No activation was seen on cotransfection of GAL4-RAP74 and VP16-RAP74, or of GAL4-RAP30 and VP16-RAP30 (Fig. 3b), or of nonspecific fusion proteins like VP16-R6K (ref. 14) with GAL4-RAP74. But cotransfection either of GAL4-RAP74 and VP16-RAP30 or of GAL4-RAP30 and VP16-RAP74 produced a significant augmentation of CAT activity. These results indicate that RAP30 and RAP74 probably do not form homodimers *in vivo* but that RAP30 and RAP74 do complex *in vivo*, either directly or indirectly through some other protein(s) or factor.

RAP74 (amino acid residues 1–207) also interacted strongly with RAP30 (data not shown) by this same assay, suggesting

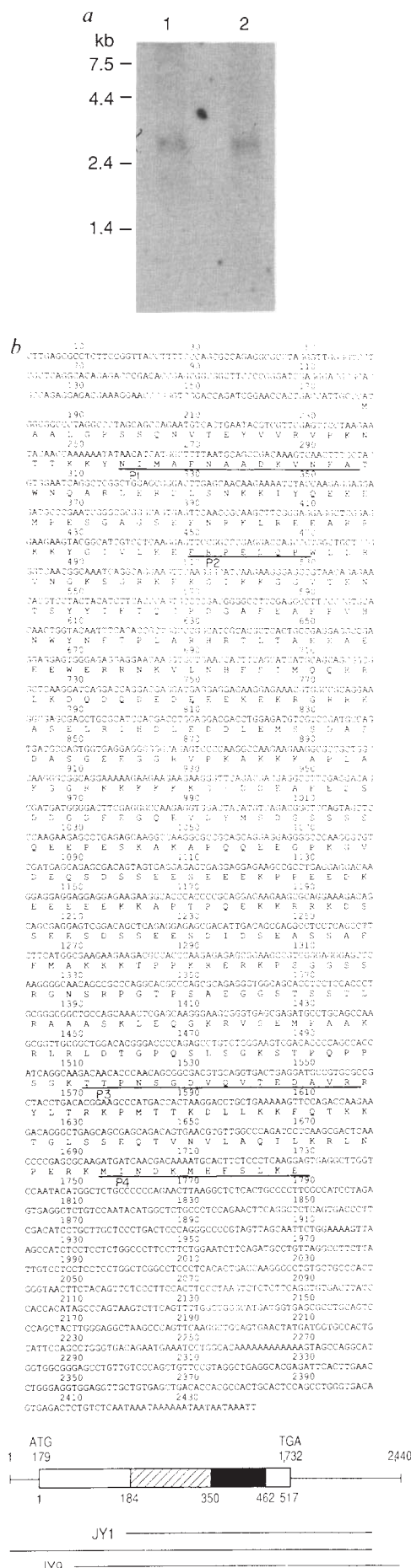
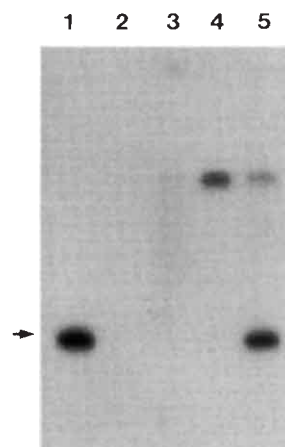


FIG. 2 Transcription activation by RAP74 produced in bacteria. *In vitro* transcription activity of recombinant subunits of the TFIIIF was assayed as reported before<sup>1</sup> and RNA transcripts were analysed on 6% polyacrylamide 7M urea gel. The reaction was done in a mixture of 30  $\mu$ l containing 0.2  $\mu$ g FA (CM-HPLC), 0.7  $\mu$ g FB (Sp), 5 ng recombinant TFIIID (provided by A. Berk), FE (S-300), 0.6  $\mu$ g FF (DEAE-Toyo), 0.5  $\mu$ g calf thymus RNA polymerase II supplemented with TFIIIF or its recombinant subunits. pMLC2AT (provided by R. Roeder) was used as a template in these reaction mixtures. For a description and nomenclature of the preparation of transcription factors, see ref. 1. Lanes 1, HeLa TFIIIF; 2, no addition; 3, 1.5  $\mu$ l (~10 ng) purified RAP30; 4, 3  $\mu$ l (~30 ng) purified RAP74; 5, mixture of 1.5  $\mu$ l RAP30 and 3  $\mu$ l RAP74.

**METHODS.** The cDNAs encoding RAP30 and RAP74 were cloned into the *Nco*I and *Bam*HI sites of pET3d (provided by F.W. Studier). *Escherichia coli* strain BL21 (DE3, Novagene) containing each clone or the parental plasmid was cultured in LB media containing ampicillin (100  $\mu$ g ml<sup>-1</sup>) at 37 °C. Cells were induced with IPTG (1 mM) when the optical density at 600 nm (OD<sub>600</sub>) reached 0.6. After 3 h, the bacterial cells were collected by centrifugation at 4,000g for 10 min. The pellet was resuspended in 1/20 of the original culture volume using 20 mM Tris-HCl pH 8.0, 5 mM EDTA, and 25 mM NaCl containing 0.5 mM PMSF and 10 mM  $\beta$ -mercaptoethanol. After freezing and thawing, cells were broken by sonication on ice, and centrifuged at 15,000g for 10 min. For RAP30, the pellet was extracted with 6 M guanidine-HCl, 50 mM Tris-HCl pH 8.0 and centrifuged at 15,000g for 10 min. The resultant supernatant was rapidly dialysed against 20 mM Tris-HCl pH 8.0, 0.2 mM EDTA, 0.3 M NaCl with several changes of buffer. After centrifugation at 15,000g for 10 min, the supernatant was dialysed against buffer B (20 mM Tris, pH 8.0, 0.2 mM EDTA, 20% Glycerol) containing 0.1 M NaCl, and 10 mM



$\beta$ -mercaptoethanol. For RAP74, the supernatant of the cell sonicate was directly applied onto a phosphocellulose column (P11, Whatman) equilibrated with buffer B, and recombinant RAP74 was stepwise-eluted with buffer B containing 1.0 M NaCl after washing the column with buffer B containing 0.5 M NaCl. The fraction was dialysed against buffer B containing 0.1 M NaCl.

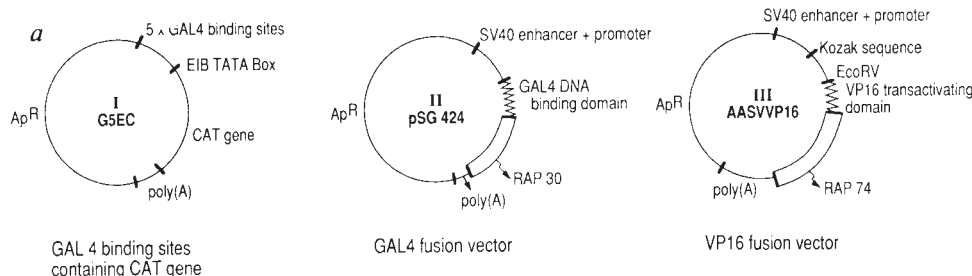
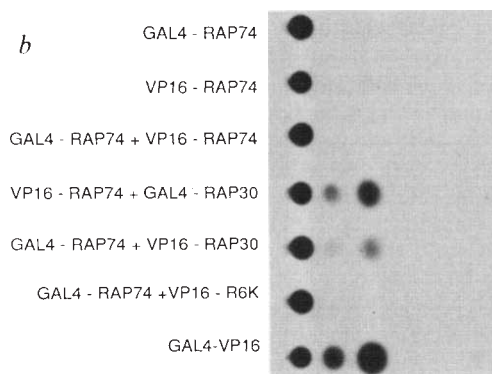


FIG. 3 CAT assay and plasmids used in the detection of RAP74-RAP30 interaction. *a*, The three plasmids used are schematically represented. Plasmid I, G5EC, contains five binding sites for the GAL4 protein and the E1B TATA box in front of the CAT gene. Plasmid II contains the sequence encoding the DNA-binding domain of the yeast GAL4 protein (amino-acid residues 1-147) downstream of the SV40 promoter and enhancer. Both of these plasmids and the plasmid producing a GAL4-VP16 fusion protein were provided by I. Sadowski and M. Ptashne<sup>13,17</sup>. A full-length cDNA clone of RAP30 was obtained by polymerase chain reaction-mediated amplification of a  $\lambda$ gt10 library derived from HeLa cells. Two primers (100 ng of primers and 1  $\mu$ g of cDNA library) were used, 5'-GAATTCGCCATGGCCGAACGTGGG-GAACTC and 5'-GCGGCCGACGTCGACCCCGTCACTCTTTCTTCTCTCC, which are specific for the sense and antisense ends of RAP30 cDNA<sup>6</sup>. A 750-bp fragment was amplified and its sequence agreed with the published sequence of RAP30 (ref. 6), except that there was an extra adenine between residues 794 and 795. This was found in numerous isolates and we consider it a correction to the published sequence. While this manuscript was being reviewed, a report describing a similar correction appeared<sup>18</sup>. The cDNA fragment was digested with *Eco*RI and *Not*I and cloned into the *Eco*RI-*Not*I site of plasmid Bluescript KS (Stratagene). The cDNA for RAP74 or RAP30 was inserted downstream of the GAL4 sequence to generate GAL4-RAP74 or GAL4-RAP30 fusion proteins. Plasmid III contains the sequence encoding the VP16 transactivating domain (amino-acid residues 413-490) downstream of the SV40 promoter and enhancer, and an optimal mammalian translation initiation sequence<sup>14</sup>. Immediately downstream of the VP16 sequence, the cDNA for RAP74, RAP30 or R6K was inserted. All of these plasmids were cotransfected with G5EC, either individually or in combination. Details of the construction of these plasmids are available on request from the authors. *b*, CV1 (African green monkey kidney) cells were obtained from



American Type culture collection and were grown in DMEM supplemented with 10% fetal calf serum, 10 units penicillin per ml<sup>-1</sup> and 10  $\mu$ g streptomycin ml<sup>-1</sup>. The reporter and effector plasmids (10  $\mu$ g each) were cotransfected by calcium phosphate precipitation into roughly 10<sup>6</sup> CV1 cells (in 10-cm plates). Precipitates were left on cells for 5-6 h and then the cells were treated for 3 min with 15% glycerol in DMEM. Cells were washed twice with PBS before the addition of fresh growth media. The cells were collected 48 h post-transfection; extracts were made by freezing and thawing five times and assayed for CAT activity as previously described<sup>19</sup>.



that the highly charged region of RAP74 is probably not involved in this interaction. Our results suggest additional roles for the cluster of charged amino acids beyond their action in known transcription activators. □

Received 5 September; accepted 7 November 1991.

1. Kitajima, S., Kawaguchi, T., Yasukochi, Y. & Weissman, S. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6106–6110 (1989).
2. Inostroza, J., Flores, O. & Reinberg, D. *J. biol. Chem.* **266**, 9304–9308 (1991).
3. Sumimoto, H., Ohkuma, Y., Yamamoto, T., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9158–9162 (1990).
4. Flores, O., Ha, I. & Reinberg, D. *J. biol. Chem.* **265**, 5629–5634 (1990).
5. Kitajima, S. *et al. Nucleic Acids Res.* **18**, 4843–4849 (1990).
6. Söpta, M., Burton, Z. F. & Greenblatt, J. *Nature* **341**, 410–414 (1989).
7. Conaway, J. W. & Conaway, R. C. *Science* **248**, 1550–1553 (1990).
8. Burtowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56**, 549–561 (1989).
9. Conaway, R. C. & Conaway, J. W. *J. biol. Chem.* **265**, 7559–7563 (1990).
10. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. *Genes Dev.* **2**, 718–729 (1988).
11. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185**, 60–89 (1990).
12. Möller, W. & Amons, R. *FEBS Lett.* **186**, 1–7 (1985).
13. Sadowski, I. & Ptashne, M. *Nucleic Acids Res.* **17**, 7539 (1989).
14. Vasavada, H., Ganguly, S., Germino, F. J., Wang, Z. X. & Weissman, S. M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
15. Lathe, R. *J. molec. Biol.* **183**, 1–12 (1985).
16. Swaroop, A. & Weissman, S. M. *Nucleic Acids Res.* **16**, 8739 (1988).
17. Sadowski, I., Ma, J., Triezenberg, S. J. & Ptashne, M. *Nature* **335**, 563–564 (1988).
18. Horikoshi, M., Fujita, H., Wang, J., Takada, R. & Roeder, R. G. *Nucleic Acids Res.* **19**, 5436 (1991).
19. Ganguly, S., Vasavada, H. & Weissman, S. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5247–5251 (1989).

ACKNOWLEDGEMENTS. We thank K. Williams and C. Stone for the peptide sequence analysis, W.-J. Xu for technical assistance and M. Weiler for manuscript preparation. This work was supported by an NCI Outstanding Investigator Grant, and by a grant from the International Human Frontier Science Program Organization.

## A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II

Ann Finkelstein, Corwin F. Kostrub, Joyce Li\*, David P. Chavez, Bo Qing Wang, Shi Min Fang, Jack Greenblatt\* & Zachary F. Burton†

Department of Biochemistry and Agricultural Experiment Station, Michigan State University, E. Lansing, Michigan 48824, USA

\* Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5G 1L6, Canada

**RAP30/74 (also known as TFIIF (refs 1, 2),  $\beta\gamma$  (ref. 3) and FC (ref. 4)) is one of several general factors required for initiation by RNA polymerase II (reviewed in refs 5–7). The small RAP30 subunit of RAP30/74 binds directly to polymerase and appears structurally and functionally homologous to bacterial sigma factors in their RNA polymerase-binding region<sup>8–10</sup>. RAP30/74 or recombinant RAP30 suppresses nonspecific binding of RNA polymerase II to DNA (refs 10, 11) and is required for RNA polymerase II to assemble stably into a preinitiation complex containing promoter DNA and the general factors TFIID, TFIIA and TFIIB (refs 2, 12, 13); both RAP30 and RAP74 are physical components of the preinitiation complex<sup>2</sup>. A complementary DNA encoding human RAP30 has been isolated<sup>8</sup>, and here we report the isolation of a cDNA encoding human RAP74. RAP30 and RAP74 produced in *Escherichia coli* can be used in place of natural human RAP30/74 to direct accurate transcription initiation by RNA polymerase II *in vitro*.**

Human RAP74 purified by affinity chromatography on an RNA polymerase II column<sup>14–16</sup> and separated from other proteins by reversed-phase chromatography was digested with

trypsin or endo Asp-N protease. Isolated RAP74 peptides were sequenced. Two oligonucleotide probes were designed from these sequence data, and rabbit antibody was generated against the RAP74 sequence DLSNKKIYQ (single-letter amino-acid code). Using these probes, five independent human cDNA clones were identified, all derived from the same gene. Figure 1 shows the nucleotide sequence and derived protein sequence of one clone likely to be a full-length cDNA. The subunit relative molecular mass of RAP74 is 58,254 ( $M_r$  58.254K), much smaller than the apparent  $M_r$  of 74K determined by denaturing gel electrophoresis (SDS-PAGE). Recombinant RAP74 produced using this clone, however, migrates with the same electrophoretic mobility as human RAP74 (data not shown).

The evidence that this cDNA encodes human RAP74 is compelling. First, each amino-acid sequence determined for RAP74 is in the coding frame (Fig. 1). Second, antibody against DLSNKKIYQ (positions 50–58) reacts with authentic RAP74. This antibody recognizes a 74K human protein that binds to RNA polymerase II (Fig. 2b, lane 4), just as anti-RAP30 antibody recognizes a 30K protein that binds to polymerase (Fig. 2a, lane 3). Moreover, both RAP30 (Fig. 2a, lane 2) and RAP74 (Fig. 2b, lane 2) were specifically immunoprecipitated from HeLa cell nuclear extracts using anti-RAP30 antibodies and were detected in western blots using anti-RAP30 and anti-DLSNKKIYQ antibody, respectively. Neither RAP30 or RAP74 was immunoprecipitated using anti- $\beta$ -galactosidase antibody (Fig. 2a, lane 1; Fig. 2b, lane 1).

Inspection of the amino-acid sequence suggests RAP74 has a globular N-terminal domain (residues 1–179), a charged central domain (residues 180–356), and a globular C-terminal domain (residues 357–517) (Fig. 3). RAP74 is a highly charged protein (38% K + R + D + E). Much of the charge is concentrated in the central domain, which is characterized by clusters of acidic residues with interspersed smaller clusters of basic residues (55% K + R + D + E, 35% D + E, 20% K + R). RAP74 lacks recognizable DNA-binding motifs, such as helix-turn-helix, zinc-finger or B-zip. Searches of recent updates of the GenBank, PIR and SwissProt databases failed to identify proteins highly homologous to RAP74.

ATP hydrolysis between the  $\beta$  and  $\gamma$  phosphates is required for accurate initiation of transcription by RNA polymerase II (ref. 17). Hydrolysis of ATP precedes, but is tightly coupled to, formation of the first phosphodiester bond in the RNA chain<sup>18</sup>. Because RAP74 may bind ATP<sup>8</sup> and have a role in ATP hydrolysis in initiation, RAP74 was searched for nucleotide-binding sequences. The best match to a nucleotide-binding sequence, GPQSLSGKST (positions 428–437), is very similar to the phosphate binding loop (P-loop) of human thymidine kinase (GPMFSGKST)<sup>19</sup>; however, an extra amino acid inserted into the consensus P-loop sequence GXXXXGKS<sup>19</sup> suggests it may not be functional. The putative RAP74 P-loop is repeated imperfectly between positions 438 and 446 (PQPPSGKTT), as if RAP74 may have two consecutive P-loops. Clearly, better characterization of the ATP-binding activity of RAP74 and mutagenesis will be required to assess critically the importance of this sequence.

DNA strand separation at the promoter should be tightly coupled to initiation. An ATP-dependent DNA helicase activity copurified with human RAP30 (ref. 8); this activity may be intrinsic to RAP30, RAP74 or another polypeptide that associates with RAP30/74. Highly purified TFIIF (ref. 20) and FC (ref. 4) lacked helicase activity. Using an *E. coli* gene-expression system, we have produced recombinant human RAP30 and RAP74. Recombinant RAP74 did not seem to have intrinsic adenylation or helicase activity, either in the absence or presence of recombinant RAP30 (data not shown), suggesting an additional factor(s) is necessary for activity (such as another general initiation factor). Recombinant RAP30 and RAP74 were, however, active in initiation of transcription (Fig. 4). This was shown by adding them, separately or in combination, to an

† To whom correspondence should be addressed.

extract of HeLa cell nuclei depleted of RAP30/74 by immunoprecipitation<sup>16</sup>.

The adenovirus major late promoter used in these assays produces a 536-nucleotide runoff transcript. Control extract prepared with anti- $\beta$ -galactosidase antibody was active for accurate transcriptional activity<sup>16</sup> (Fig. 4, lane 1). Somewhat

surprisingly, RAP74 stimulated transcription when added to the control extract (Fig. 4, lanes 3 and 4). RAP30/74 was not limiting for transcription in HeLa cell nuclear extracts<sup>16</sup>, and consistent with this, RAP30 was not limiting for transcription (compare lanes 1 and 2). Because some HeLa cell nuclear extracts seem to be relatively deficient in RAP74 (ref. 14), perhaps RAP74

```

GAATTCGCGCCTCTGGCGCTGGGAAAAGTAACCCGGTCCGACCAGATCGGAACCACT 60
GACCATTGCCATGGCGGCCCTAGGCCCTAGCAGCCAGAATGTCACCTGAATACGTCGTTT 120
      M A A L G P S S Q N V T E Y V V R 17
GAGTTCCTAAGAATACAACCAAAAAATATAACATCATGGCTTTTAAATGACGCCGACAAAG 180
      V P K N T T K K Y N I M A F N A A D K V 37
-----
TCAACTTTGTCTACGTGGAATCAGGCTCGGCTGGAGCGGGACTTGAGCAACAAGAAATCT 240
      N F A T W N Q A R L E R D L S N K K I Y 57
-----
ACCAAGAGGAGGAGATGCCGAATCGGGCGGGCAGTGAGTTCACCGCAAGCTTCGGG 300
      Q E E E M P E S G A G S E F N R K L R E 77
-----
AGGAGGCTCGGAGGAAGAAGTACGGCATCTCTCAAGGAGTTCGGGCCCGAGGACCAGC 360
      E A R R K K Y G I V L K E F R P E D Q P 97
-----
CCTGGCTGCTCGGGTCAACGGCAATCAGGAGGAAGTTCAGGGCATCAAGAGGGAG 420
      G L L R V N G K S G R K F K G I K K G G 117
-----
GCGTAACAGAGAACACGCTCTACTACATCTTACCCAGTGCCTCGGCGGCTTCGAGG 480
      V T E N T S Y Y I F T Q C P D G A F E A 137
-----
CCITCCCCGTGCACAACTGGTACAATTTACACCGCTGGCCCGGCTCGCACCTCTCACT 540
      F P V H N W Y N F T P L A R H R T L T A 157
-----
CCGAGGAGGCGGAGGAGGAGTGGGAGGAGGAACAAGGTGCTGAACCACTTCAGCATCA 600
      E E A E E W E R R N K V L N H F S I M 177
-----
TGACGAGCGGGCGGCTCAAGGATCAGGACGAGGAGGAGTGGAGGAGGAGGAGAAAC 660
      Q Q R R L K D Q D Q D E D E E E K E K R 197
-----
GTGGCGCAGGAAGGCGAGGAGCTGCGCATCCACGACCTGGAGGACGACCTGGAGATGT 720
      G R R K A S E L R I H D L E D D L E M S 217
-----
CGTCCGATGCCAGTGTGCGAGTGGTGGAGGAGGGGGCAGAATCCCAAGGCCAAGAAGA 780
      S D A S D A S G E E G G R I P K A K K K 237
-----
AGGCGCGCTGGCCCAAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
      A P L A K G G R K K K K K K G S D D E A 257
-----
CCITCGAGGACGAGTGTGGGAGCTTCGAGGCGCAAGAGGTGGAGTACATGTCAGAG 900
      F E D S D D D G D E L R I E G G Q E V D Y M S D G 277
-----
GCTCCAGTAGCTCCCAAGAGAGCTGAGAGCAAGGCCAGCGCGCAGCAGGAGGAGG 960
      S S S S Q E E P E S K A K A P Q Q E E G 297
-----
GGCCCAAGGGTGTGATGAGCAGAGCAGCAGTGTAGGAGAGTGTAGGAGGAGGAGGAGG 1020
      P K G V D E Q S D S S E E S E E E K P P 317
-----
CTGAGGAGGACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
      E E D K E E E E K K A P T P Q E A K K R 337
-----
GCAGGAAGACAGCAGCAGGAGTGGAGCAGCTCAGAGGAGAGGAGGAGGAGGAGGAGG 1140
      R K D S S E E S D S S E E S D I D S E A 357
-----
CCTCTCAGCCCTCTTCTATGGCGAAGAGAGAGCAGCCCAAGAGAGAGGAGGAGGAGG 1200
      S S A L F M A K K K T P P K R E R K P S 377
-----
CGGGAGGAGCTCAAGGGGCAACAGCGCCCGGCGCAGCCAGGAGGAGGAGGAGGAGGAG 1260
      G G S S R G N S A R L D T G P Q S L S G K S T 397
-----
CCTCTCCACCTGCGGGCGGCTGCCAGCAAACTCGAGCAAGGAGGAGGAGGAGGAGGAG 1320
      S S T L R A A A S K L E Q G K R V S E M 417
-----
TGCTCGAGCAGGAGGCTGCGGCTGGACACGGGAGCCAGAGCCTGTCTGGGAAGTCGA 1380
      P A K R L R L D T G P Q S L S G K S T 437
-----
CACCOCAGCCACCATCAGGCAAGCAACACCCACAGCGGCGAGCTGCAGGTGACTGAGG 1440
      P Q P P S G K T T P N S G D V Q V T E D 457
-----
ATGCCGTGCGCGCTACCTGACACGGAAGCCATGACCACTAAGGACCTGCTGAAAAAGT 1500
      A V R R Y L T R K P M T T K D L L K K F 477
-----
TCCAGACCAAGAGCAGGGCTGAGCAGCGAGCAGAGTGAAGTGTGGCCAGATCC 1560
      Q T K K T G L S S E Q T V N V L A Q I L 497
-----
TCAAGCGACTCAACCCGAGCGCAAGATGATCAACGACAAAATGCACTTCTCCCTCAAG 1620
      K R L N P E R K M I N D K M H F S L K E 517
-----
AGTGAGGCTTGGTCCAATACATGGCTCTGCCCGCCAGAACTTAAGGCTCTCACTGCCCT 1680
      TCGCCATCTAGAGTGAAGGCTGTGTCATACATGGCTCTACCTCCAGAACTCAGGCT 1740
      CTCAGTGACCTTCGACATCTGCTTGTCTGCTGACTCCAGGGCCCGCTAGTTAGCAAT 1800
      TGTGAAAGTAAAGGCATCTCTCTCTGCGCTCTCTCTGGAATCTTCAGATGCCCTG 1860
      TTAGGCTCTTATTGTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
      TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
      GCTGTGACTTATCCACCATAGCCAGTAAGTCTTCACTTTTGGCTGGGCATGATGGTG 2040
      AGCGCTGCACTCCAGCTACTTGGGAGGCTAAGCCAGTTCAAGGCTGCACTGAACATAT 2100
      GATGTGCACTGCACTTCAAGCTGGTGACAGAAATCCTGGCAAAAAAAGGAGGAGG 2160
      AAAAGTACGAGGAGGAGTGGTGGGAGGCTGTTGCTCCAGCTGTTCCGTAGGCTGAGG 2220
      CGAGATCACTTGAACCTGGGAGGTGGAGGTTGCTGTGAGCTGACACCAAGCCACTGAC 2280
      TCCAGCTGGTGTGACAGTGTCTCTCAATAAATAAAAAATAATAAAAAA 2340
      AAAAAAAAAA

```

FIG. 1 The nucleotide sequence of a cDNA encoding human RAP74. The derived amino-acid sequence of RAP74 is shown beneath the nucleotide sequence. Coordinates for the nucleotide and amino-acid sequences are shown to the right of each line. Amino-acid sequences determined by Edman degradation are underlined with a dashed line. Just upstream of the poly(A) 3' tail of the cDNA are three consensus polyadenylation signals (AAUAAA), between positions 2,312 and 2,334. Potential nucleotide-binding sequences are located between amino-acid positions 428 and 446 (see text for discussion). The sequence DSSEES is repeated three times in RAP74.

**METHODS.** RAP74 was digested either with trypsin or endo Asp-N, and peptides were isolated on a 1 mm  $\times$  5 cm C-8 column (RP300; Applied Biosystems) in 0.1% trifluoroacetic acid buffer, using 90% acetonitrile in 0.85% trifluoroacetic acid as the eluting solvent. Peptide sequence was determined using an Applied Biosystems 477A Automated Protein/Peptide Sequencer in the Michigan State University Macromolecular Structure, Sequencing and Synthesis Facility. A  $\lambda$ gt11 library constructed from cDNA derived from human endothelial cells (gift of V. Dixit) was screened with oligonucleotide probes. Nine clones were isolated from  $3 \times 10^5$  plaques that screened positive in duplicate with the estimated oligomer GATAAAGTGA-ACTTTGCTACATGGAAYCARGC (derived from the sequence DKVNFATWNQA) using a sodium chloride/sodium citrate hybridization protocol at 53 °C. For screens with antibody against DLSNKKIYQ peptide, a  $\lambda$  ZAP library constructed with cDNA derived from human HeLa cells (the library was the gift of D. Reinberg) was used. Twenty clones were isolated from a screen of  $3 \times 10^6$  plaques that tested positive with anti-peptide antibody. Phage from both the oligonucleotide and antibody screens were then tested for hybridization with the probe YTGRTA(A/G/T)ATYTTYTTRTT (derived from the sequence NKKIYQ) in tetramethyl ammonium chloride. Four independent cDNAs obtained from the oligonucleotide screen and one cDNA from the antibody screen tested positive with this second probe. Sequence analysis showed that all of these clones encode human RAP74. The one apparently full-length clone that we obtained was isolated from the endothelial cell library. DNA sequence determination was by the method of Sanger (Sequenase, US Biochemicals). DNA and amino-acid sequence analysis was done using the University of Wisconsin Genetics Computer Group software package<sup>28</sup>.



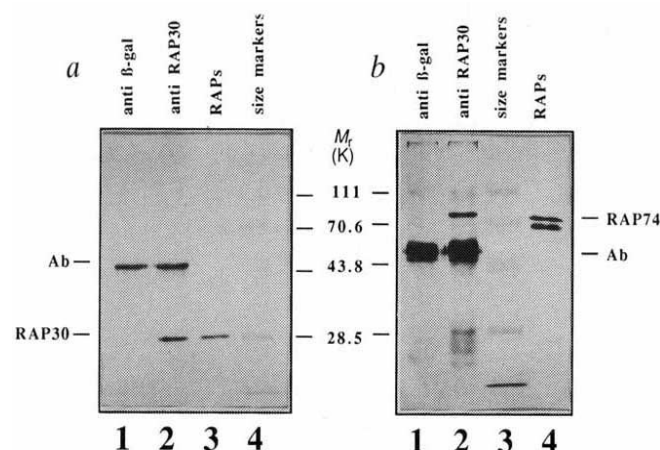


FIG. 2 Immunoprecipitation of RAP30/74. *a*, Western blot developed with anti-RAP30 antibody. *b*, Western blot developed with anti-DLSNKKIYQ antibody. Samples were immunoprecipitated from human HeLa extracts with anti- $\beta$ -galactosidase (anti  $\beta$ -gal) antibody (lane 1) or anti-RAP30 antibody (lane 2). A RAP fraction (RAPs) was the salt eluate of an RNA polymerase II column that had been loaded with HeLa cell extract<sup>14-16</sup>; roughly 100 ng RAP30 and 50 ng RAP74 are present in these lanes. Antibody heavy chain from the immunoprecipitations is indicated on the blots (Ab). Prestained  $M_r$  markers were from Bethesda Research Laboratories.

**METHODS.** The method for immunoprecipitation was as used for constructing a RAP30/74-depleted extract<sup>16</sup> (Fig. 4). Two samples of an extract of HeLa cell nuclei (1 ml) were suspended in 20 mM HEPES pH 7.9, 20% glycerol, 0.5 M NaCl, 0.2 mM EDTA, 0.2 mM EGTA, 0.1 mM DTT. These extracts were cleared of insoluble material by centrifugation. Affinity-purified anti-RAP30 (80  $\mu$ g) was added to one sample (lane 2) and 80  $\mu$ g affinity purified anti- $\beta$  galactosidase was added to the other (lane 1). Extracts were mixed with antibody for 1 hr at 4 °C. Six protein A-Sepharose (Pharmacia) columns (0.5 ml volume) were equilibrated with the same buffer containing 1 mg ml<sup>-1</sup> BSA (Sigma; fraction V). Each extract was passed three times through a 0.5-ml protein A-Sepharose column. Flow-through fractions were collected and a second addition of the appropriate antibody was made to each sample as described above. The extracts were each passed three times through a second protein A-Sepharose column. These flow-through fractions were then each cycled through a third protein A-Sepharose column to remove any remaining traces of antibody. These final flow through fractions were concentrated in Amicon Centricon 10 Microconcentrators, preadsorbed with a 1 mg ml<sup>-1</sup> solution of BSA, and extracts were dialysed into the above buffer containing 0.1 M KCl, in place of NaCl, and 2 mM dithiothreitol. The extract prepared with anti-RAP30 antibodies is referred to as the RAP30/74-depleted extract (RAP30/74 DE); the extract prepared with anti- $\beta$  galactosidase antibodies is referred to as the control extract (CE). The first protein A-Sepharose column from each extract preparation was washed with 3 ml of the column buffer, and SDS-PAGE sample buffer (1 ml) was cycled through the columns three times, to elute proteins associated with antibodies bound to the column. This material (30  $\mu$ l) was loaded into the indicated wells of an SDS-PAGE gel. Electrophoresis and electrotransfer were according to standard procedures. Western blot detection was with first antibody: affinity-purified chicken anti-RAP30 antibodies; and second antibody: rabbit anti-chicken alkaline phosphatase conjugate (Pel-Freeze) (*a*); and first antibody: affinity-purified rabbit anti-DLSNKKIYQ antibody; and second antibody: goat anti-rabbit alkaline phosphatase conjugate (Bio Rad) (*b*). Colour development was with standard reagents according to the manufacturer's instructions (Bio Rad).

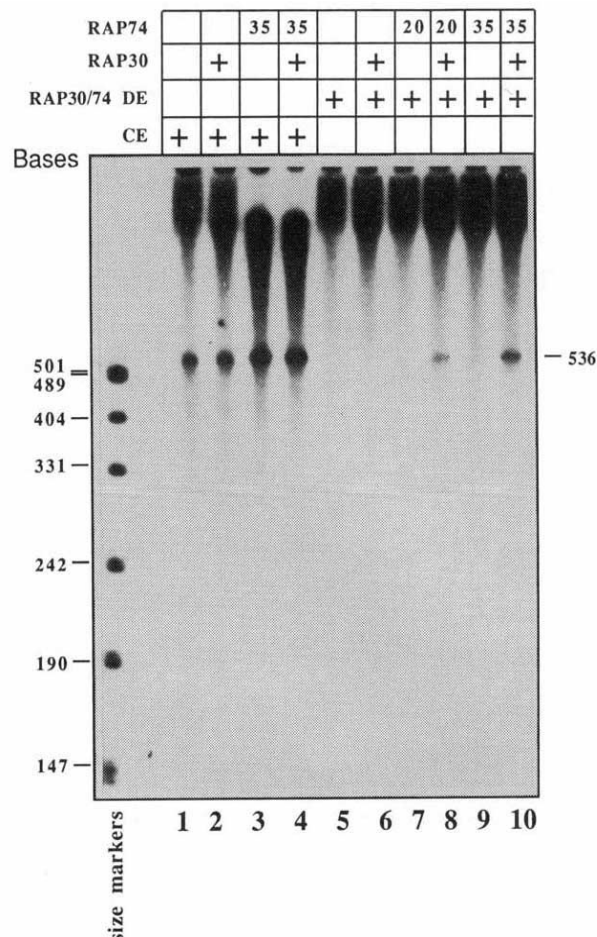


FIG. 4 RAP30 and RAP74 produced in *E. coli* reconstitute accurate transcription *in vitro*. The 536-nucleotide runoff transcript from the adenovirus major late promoter is indicated. Additions to a control extract (CE) and a RAP30/74-depleted extract (RAP30/74 DE) are indicated at the top of the figure. Recombinant RAP74 (20 or 35 ng) was added to reactions where indicated. Each addition of RAP30 was 100 ng.

**METHODS.** Preparation of the RAP30/74 DE and CE extracts is described in the legend to Fig. 2. Recombinant RAP30 and RAP74 were produced using the *E. coli* expression vector pET11d developed by Studier and colleagues (Novagen). This vector puts cloned DNA under control of the strong bacteriophage T7 promoter. Clones and methods for RAP30 and RAP74 production and purification will be provided on request. Methods for *in vitro* transcription were essentially as described previously<sup>16</sup>. The template was pMLP digested with *Sma*I restriction endonuclease. Extracts with the indicated additions were pre-incubated for 1 h at 30 °C. Nucleotide triphosphates were added to a final concentration of 600  $\mu$ M ATP, CTP, UTP, 25  $\mu$ M GTP, and 5  $\mu$ M  $[\alpha\text{-}^{32}\text{P}]\text{GTP}$  was added to each reaction. After 30 min, reactions were stopped and radioactive RNA was isolated and analysed. Size markers are pUC19 DNA digested with *Hpa*II and 3' end-labelled with  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$  and Klenow DNA polymerase I.

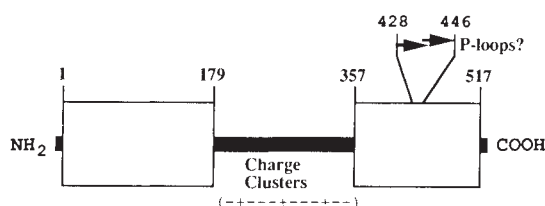


FIG. 3 Proposed domain structure for human RAP74 (see text for discussion).

becomes inactivated during initiation (for example by phosphorylation<sup>14</sup> or adenylation<sup>8</sup>) and must be regenerated for new rounds of transcription to occur.

The RAP30/74-depleted extract was inactive for accurate transcriptional initiation<sup>16</sup> (lane 5). Addition of RAP30 alone did not stimulate transcription (lane 6), nor did addition of RAP74 alone (lanes 7 and 9). Addition of both RAP30 and RAP74 strongly stimulated accurate transcriptional activity (lanes 8 and 10), and increasing the quantity of RAP74 increased the yield of transcript (compare lanes 8 and 10). We conclude that both RAP30 and RAP74 are required for accurate transcription by RNA polymerase II. As the RAP30 and RAP74 used in this experiment were produced in *E. coli*, the clones we have isolated encode the active human transcription factors. The use

of recombinant human RAP30 and RAP74, as well as recombinant TFIID (refs 21–23), TFIIB (ref. 24) and TFIIE (refs 25–27) will contribute to establishment of a defined system for accurate initiation by human RNA polymerase II *in vitro*. □

Received 30 September; accepted 15 November 1991.

1. Flores, O., Maldonado, E., Burton, Z. F., Greenblatt, J. & Reinberg, D. *J. Biol. Chem.* **263**, 10812–10816 (1988).
2. Flores, O. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 9999–10003 (1991).
3. Conaway, J. W. & Conaway, R. C. *J. Biol. Chem.* **264**, 2357–2362 (1989).
4. Katayama, S. *et al. Nucleic Acids Res.* **18**, 4843–4849.
5. Saltzman, A. R. & Weinmann, R. *FASEB J.* **3**, 1723–1733 (1989).
6. Mermelstein, F. H., Flores, O. & Reinberg, D. *Biochim. biophys. Acta* **1009**, 1–10 (1989).
7. Sawadogo, M. & Sentenac, A. *Rev. Biochem.* **59**, 711–754 (1990).
8. Sotola, M., Burton, Z. F. & Greenblatt, J. *Nature* **341**, 410–414 (1989).
9. McCracken, S. & Greenblatt, J. *Science* **253**, 900–902 (1991).
10. Killeen, M. T. & Greenblatt, J. *Molec. cell Biol.* (in the press).
11. Conaway, J. W. & Conaway, R. C. *Science* **248**, 1550–1553 (1990).
12. Conaway, R. C., Garrett, K. P., Hanley, J. P. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6205–6209 (1991).
13. Buratowski, S., Sotola, M., Greenblatt, J. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7509–7513 (1991).
14. Sotola, M., Carthew, R. W. & Greenblatt, J. *J. Biol. Chem.* **260**, 10353–10361 (1985).
15. Burton, Z. F., Ortalan, L. G. & Greenblatt, J. *EMBO J.* **5**, 2923–2930 (1986).
16. Burton, Z. F., Killeen, M., Sotola, M., Ortalan, L. G. & Greenblatt, J. *Molec. cell Biol.* **8**, 1602–1613 (1988).
17. Bunick, D., Zandomeni, R., Ackerman, S. & Weinmann, R. *Cell* **29**, 877–886 (1982).
18. Conaway, R. C. & Conaway, J. W. *J. Biol. Chem.* **263**, 2962–2968 (1988).
19. Saraste, M., Sibbald, P. R. & Wittinghofer, A. *Trends Biochem. Sci.* **15**, 430–434 (1990).
20. Flores, O., Ha, I. & Reinberg, D. *J. Biol. Chem.* **265**, 5629–5634 (1990).
21. Kao, C. C. *et al. Science* **248**, 1646–1649 (1990).
22. Peterson, M. G., Tanese, N., Pugh, B. F. & Tjian, R. *Science* **248**, 1625–1630 (1990).
23. Hoffmann, A. *et al. Nature* **346**, 387–389 (1990).
24. Ha, I., Lane, W. S. & Reinberg, D. *Nature* **352**, 689–695 (1991).
25. Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M. & Roeder, R. G. *Nature* **354**, 398–401 (1991).
26. Sumimoto, H., Ohkuma, Y., Sinn, E., Shimasaki, S., Kato, H., Horikoshi, M. & Roeder, R. G. *Nature* **354**, 401–404 (1991).
27. Peterson, M. G., Inostroza, J., Maxon, M. E., Flores, O., Admon, A., Reinberg, D. & Tjian, R. *Nature* **354**, 369–373 (1991).
28. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).

ACKNOWLEDGEMENTS. We thank J. Leykam for help with protein sequencing and isolation. This research was supported by the American Cancer Society, the NIH, the Medical Research Council of Canada and the National Cancer Institute of Canada. Z.B. was further supported by funds from Michigan State University. J.G. is an International Research Scholar of the Howard Hughes Medical Institute.

## Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage

Andreas Meisel\*†, Thomas A. Bickle\*†, Detlev H. Krüger & Cornelia Schroeder\*§

\* Institute of Virology, Humboldt University Medical School (Charité), D-1040 Berlin, Germany

† Microbiology Department, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

**TYPE III restriction/modification enzymes recognize short, non-palindromic sequences that can be methylated on only one strand, with the paradoxical consequence that during replication of what is in effect hemimethylated DNA totally unmodified sites arise<sup>1</sup>. Why the unmodified sites are not subject to suicidal restriction was not clear. Here we show that restriction requires two unmodified recognition sites that can be separated by different distances but which must be in inverse orientation. All of the unmodified sites in newly replicated DNA are of course in the same orientation, which explains why they are not restricted. This result may be of relevance to other manifestations of anisotropy in double-stranded DNA, such as genetic imprinting<sup>2</sup>.**

The resistance of certain bacteriophages to restriction by the type III restriction/modification (R/M) system *EcoP15* was investigated. Both the genomes of coliphage T7 and its cousin T3 contain dozens of recognition sites for *EcoP15* (5'-CAGCAG-3'; Fig. 1a), but T7 is refractive whereas T3 is susceptible to restriction<sup>3</sup>. There is one striking difference between the two phage genomes: all 36 *EcoP15* recognition sites in T7 DNA are aligned unidirectionally along the DNA, whereas in T3 DNA at least two of these sites are inverted (Fig. 1b). We considered this highly significant strand bias of *EcoP15* sites in T7 DNA to be crucial for protection against restriction because only pairs of inversely oriented nonmethylated recognition sites could constitute a substrate for type III restriction endonucleases<sup>3,4</sup>. This model would explain why an *EcoP15* site in pBR322 DNA can be cleaved on the intact plasmid (which contains several sites in both orientations) but not when it is isolated in a restriction fragment<sup>5</sup>, and why complete digestion is impossible. More important, the model also explains why the DNA of cells expressing type III enzymes is not subject to suicidal restriction: inspection of Fig. 1c reveals that all unmodified sites generated during replication of modified (that is, hemimethylated) DNA will be strand-biased.

To verify our model experimentally, we constructed derivatives of M13 phage DNA with alternative arrangements of *EcoP15* sites (Fig. 2). M13 wild-type DNA contains three *EcoP15* recognition sites all with the same orientation<sup>6</sup> and is

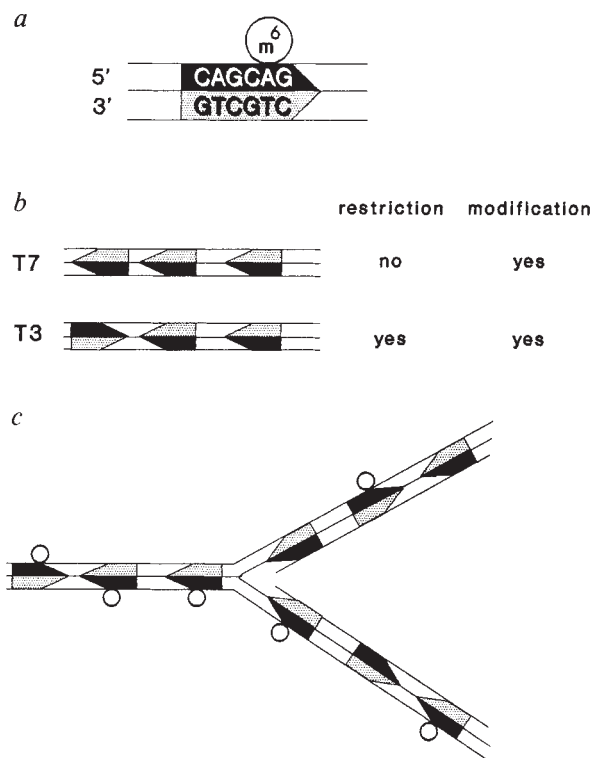


FIG. 1 Strand-bias model for DNA-*EcoP15* enzyme interaction. *a*, Recognition sequence of *EcoP15*. The arrow shows the orientation of this nonpalindromic sequence; the arrowhead indicates the direction of the cleavage site, 25–27 base pairs downstream from the recognition sequence<sup>5</sup>; m<sup>6</sup> corresponds to the methylation position<sup>10</sup>. *b*, Restriction is only possible if two unmodified *EcoP15*-sites are inversely oriented, DNA with directly repeated sites is refractory to restriction. *EcoP15* methylation is independent of the site orientation. *c*, A replication fork in *EcoP15*-coding cells. In the parental DNA molecule all *EcoP15* sites are methylated but in the two daughter molecules only those sites are modified where the methylated A residues (indicated by circles) are in the parental DNA strand. Sites with T residues in the parental strand are unmodified. All unmodified sites of a daughter molecule have the same orientation.

§ To whom correspondence should be addressed.

§ Present address: National Institute for Medical Research, Division of Virology, London NM7 1AA, UK.



resistant to restriction by *EcoP15*. Each of the constructs EM1, EM2 and EM3 has a different site inverted and all three became susceptible to cleavage by *EcoP15* *in vitro* and *in vivo*. In contrast to M13 replicative form (RF) DNA (Fig. 3a, lane 5), the circular RF DNAs of EM1, EM2 and EM3 are cleaved once by *EcoP15* (Fig. 3a, lane 6 to 8). EM1 is partially digested into smaller fragments (lane 6). Using the host strain *Escherichia coli* TG1 (ref. 7) and the *EcoP15*-coding strain TG1(P1-15hyb2cIts), we could show that EM2, unlike M13, is subject to host-controlled R/M. EM2 is restricted by a factor of 0.28 and becomes modified after passage on this strain.

We then showed that pairs of inverted nonmethylated sites are substrates for *EcoP15*. Owing to the asymmetry of the *EcoP15* site, inverted recognition sequences can be paired in two ways, namely as form I:  $\rightarrow \leftarrow$  or form II:  $\leftarrow \rightarrow$ , where one arrow is equivalent to an *EcoP15* site (Fig. 1a). This is tantamount to defining the substrate as a palindromic recognition site interrupted by a spacer of variable length. Obviously, in circular DNA, forms I and II are topologically indistinguishable. Linearizing M13 and the three constructs at their unique *AvaI* or *BamHI* sites generated eight substrate DNAs exhibiting all possible constellations of *EcoP15* sites. The following observations support the conclusion that form I is the preferred substrate of *EcoP15*. Linearization of EM1 by *AvaI* yielded a product containing two form-I pairs of sites (Fig. 2, sites 1+2 and 1+3), which were both susceptible to *EcoP15* (Fig. 3b, lane 4), whereas the *BamHI* product of EM1 that contains form-II pairs (Fig. 2, sites 3+1 and 2+1) is quite refractive to *EcoP15* (Fig. 3b, lane 12). In the two form-I pairs of EM1/*AvaI*, restriction occurred randomly at one of the corresponding sites (Fig. 3b, lane 4). Both the *BamHI* and the *AvaI* product of EM2 are substrates for *EcoP15* (Fig. 3b, lane 14); they both contain a form-I pair of sites (Fig. 2, sites 2+3). In addition, the *AvaI* product has also a form-II pair (Fig. 2, sites 1+2) and the

*BamHI* product a second form-I pair (Fig. 2, sites 2+1). But, no cleavage at site 1 can be observed. It is possible that the enzyme prefers to interact with sites 2 and 3, which are close to each other, rather than with sites 2 and 1, which are much further apart. Cutting EM3 created form II (*AvaI*) and a mixed form-I/II product (*BamHI*). Both were incompletely cleaved by *EcoP15* (Fig. 3b, lane 8, 16), although the site constellation in EM3/*BamHI* is topologically equivalent to that in EM1/*AvaI*. The close proximity of form-II sites to the form-I sites in EM3/*BamHI*, particularly considering that form I and form II share the central recognition site (Fig. 2), may impair the functional cooperation between *EcoP15* enzyme subunits located on adjacent sites.

*EcoP15* is composed of subunits encoded by the *mod* gene and *res* gene. The *mod* product alone is capable of substrate recognition and methylation whereas the *res* gene encodes the endonuclease subunit which is only active when complexed with *mod* subunits<sup>8</sup>. In the presence of the essential cofactors for methylation (AdoMet) and restriction (ATP) at physiological concentrations, methylation and restriction are competitive reactions<sup>9</sup>. Our experiments (Fig. 3a, b) were performed in the absence of AdoMet, so that only restriction can occur. Under these *in vitro* conditions a residual nicking and cleavage activity with supercoiled (Fig. 3a, lane 5) and linearized M13 RF DNA (Fig. 3b, lane 2, 10) was observed. Using physiological concentrations of the cofactors ATP (1 mM) and AdoMet (1  $\mu$ M) in the reaction assay, *BamHI*-linearized M13 RF DNA was totally resistant (Fig. 4a), whereas more than 60% of *BamHI*-linearized EM2 RF DNA was cleaved, the rest being modified

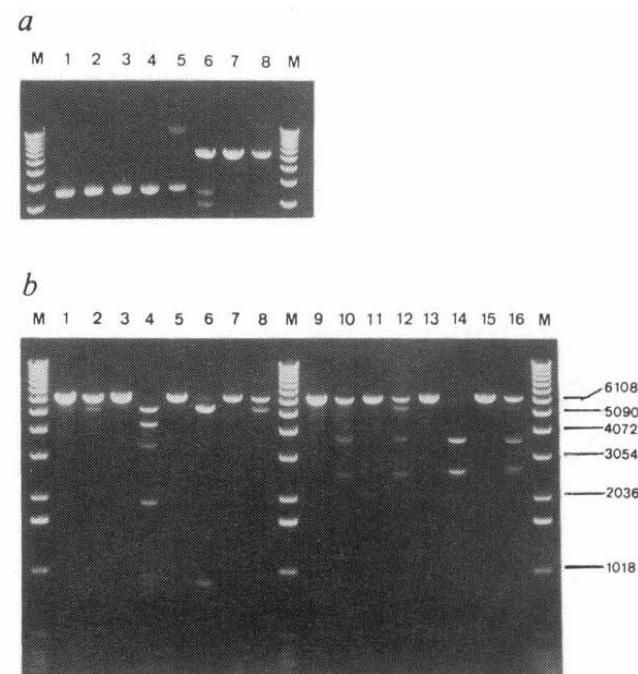
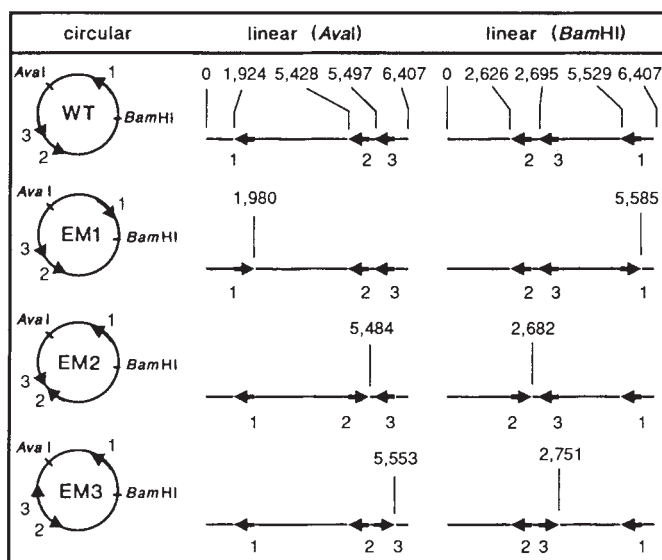


FIG. 3 *EcoP15* restriction of circular and linearized RF DNAs of M13 wild type and the mutants EM1, EM2 and EM3 *in vitro*. a, M13 (lane 1, 5), EM1 (lane 2, 6), EM2 (lane 3, 7) and EM3 (lane 4, 8) supercoiled RF DNAs without (lane 1 to 4) and with *EcoP15* treatment (lane 5 to 8). Lanes M, DNA size markers. b, RF DNAs linearized by *AvaI* (lane 1 to 8) or *BamHI* (lane 9 to 16); M13 (lane 1, 2, 9, 10), EM1 (3, 4, 11, 12), EM2 (5, 6, 13, 14) and EM3 (7, 8, 15, 16) without (odd lane numbers) or with (even lane numbers) *EcoP15* treatment. Lanes M, marker DNA fragments (sizes given to the right, in base pairs).

METHODS. Each RF DNA (300 ng) was incubated in *EcoP15* buffer<sup>8</sup> and 1 mM ATP in the absence or in the presence of 100 ng *EcoP15* for 3 h at 37 °C. Reactions were terminated by extraction with phenol followed by precipitation with ethanol. Linearization used 1  $\mu$ g RF DNA and 2 units of *AvaI* or *BamHI* in the appropriate buffer (New England BioLabs) for 2 h at 37 °C and the DNA was subsequently extracted and precipitated. Fragments were separated by electrophoresis through a 0.8% agarose gel.

FIG. 2 *EcoP15* site constellation in the RF DNA of M13 wild-type phage (WT) and of the mutants EM1, EM2 and EM3. Arrows indicate the orientations of the three *EcoP15* sites. The site constellation is represented for the circular DNA and the DNA linearized by *AvaI* or *BamHI*. Numbers above the genomes show the positions of the restriction sites and the length of the linearized DNAs (in base pairs) for M13 wild type and, if changed, for the mutants. METHODS. The mutants EM1, EM2 and EM3 were constructed by site-directed mutagenesis<sup>11</sup> and confirmed by sequencing using T7 DNA polymerase<sup>12</sup>. The oligonucleotides used to generate mutants were 30-mers with two T replacing the two A residues of the included *EcoP15* sequence. Single-strand DNA for mutagenesis and sequencing experiments and RF DNA were purified according to standard methods<sup>7</sup>.

(Fig. 4b). The amount of EM2 cleavage *in vitro* at physiological concentrations of both cofactors is in good agreement with the observed *in vivo* restriction of EM2 phage. In contrast to restriction, M13 DNA (Fig. 4a) as well as EM2 DNA (including the fragments produced by *Eco*P15 cleavage; Fig. 4b) can be fully modified.

In conclusion, *Eco*P15 restriction is a very fast reaction that

requires two unmodified recognition sites in inverted orientation. *Eco*P15 methylation is a slower reaction and is independent of the orientation of the sites. □

Received 23 September; accepted 22 October 1991.

1. Bickle, T. A. in *Nucleases* (eds Linn, S. M. & Roberts, R. J.) 85–108 (Cold Spring Harbor Laboratory Press, New York, 1983).
2. Klar, A. J. S. *Nature* **326**, 466–470 (1987).
3. Schroeder, C., Jurkschat, H., Meisel, A., Reich, J. G. & Krüger, D. H. *Gene* **45**, 77–86 (1986).
4. Krüger, D. H., Bickle, T. A., Reuter, M., Pein, C. D. & Schroeder, C. in *Nucleic Acid Methylation* (eds Clawson, G. A., Willis, D. B., Weissbach, A. & Jones, P. A.) 113–124 (Liss, New York, 1990).
5. Hadi, S. M. et al. *J. molec. Biol.* **134**, 655–666 (1979).
6. van Wezenbeek, P. M. G. F., Hulsebos, T. J. M. & Schoenmakers, J. G. G. *Gene* **11**, 129–148 (1980).
7. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
8. Hadi, S. M., Bächli, B., Iida, S. & Bickle, T. A. *J. molec. Biol.* **165**, 19–34 (1983).
9. Reiser, J. & Yuan, R. *J. biol. Chem.* **252**, 451–456 (1977).
10. Meisel, A., Krüger, D. H. & Bickle, T. A. *Nucleic Acids Res.* **19**, 3997 (1991).
11. Vandeyar, M. A., Weiner, M. P., Hutton, C. J. & Batt, C. A. *Gene* **65**, 129–133 (1988).
12. Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4076–4080 (1989).

ACKNOWLEDGEMENTS. We thank M. Gübler, P. Hübner, N. Redaschi, T. Naas (Basel) and M. Reuter (Berlin) for help with the experiments. A.M. was a recipient of an EMBL short-term fellowship. This work was supported by the Swiss National Science Foundation.

## Novel NADPH-binding domain revealed by the crystal structure of aldose reductase

J.-M. Rondeau\*†, F. Tête-Favier\*†, A. Podjarny\*,  
J.-M. Reyman†, P. Barth‡, J.-F. Biellmann‡  
& D. Moras\*§

\* Laboratoire de Cristallographie Biologique, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue R. Descartes, 67084 Strasbourg Cedex, France

† Biostructure SA, Les Algorithmes, Bâtiment Euclide, Parc d'Innovation, 67400 Illkirch-Graffenstaden, France

‡ Laboratoire de Chimie Organique Biologique, Institut de Chimie, Université Louis Pasteur, 4 rue B. Pascal, 67008 Strasbourg Cedex, France

ALDOSE reductase is the first enzyme in the polyol pathway and catalyses the NADPH-dependent reduction of D-glucose to D-sorbitol. Under normal physiological conditions aldose reductase participates in osmoregulation<sup>1</sup>, but under hyperglycaemic conditions it contributes to the onset and development of severe complications in diabetes<sup>2</sup>. Here we present the crystal structure of pig lens aldose reductase refined to an *R*-factor of 0.232 at 2.5-Å resolution. It exhibits a single domain folded in an eight-stranded parallel  $\alpha/\beta$  barrel, similar to that in triose phosphate isomerase<sup>3</sup> and a score of other enzymes. Hence, aldose reductase does not possess the expected canonical dinucleotide-binding domain<sup>4</sup>. Crystallographic analysis of the binding of 2'-monophospho-adenosine-5'-diphosphoribose, which competitively inhibits NADPH binding reveals that it binds into a cleft located at the C-terminal end of the strands of the  $\alpha/\beta$  barrel. This represents a new type of binding for nicotinamide adenine dinucleotide coenzymes.

Aldose reductase (alditol-NADP<sup>+</sup> 1-oxidoreductase; EC 1.1.1.21) belongs to a class of monomeric NADPH-dependent oxidoreductases with relative molecular masses in the range 35,000–40,000 the aldo-ketoreductases<sup>5</sup>. Within this class, sequences of aldose and aldehyde reductases are highly homologous and should therefore fold into similar three-dimensional structures<sup>6–9</sup>. Strong sequence similarities also exist with prostaglandin F synthase<sup>10</sup>, 2,5-diketo-D-gluconate reductase<sup>11</sup>, *p*-crystallin<sup>12</sup>, chlordecone reductase<sup>13</sup> and a yeast protein encoded by the *GCY* gene<sup>14</sup>. The crystal structure of aldose reductase was solved at atomic resolution (Table 1). It reveals

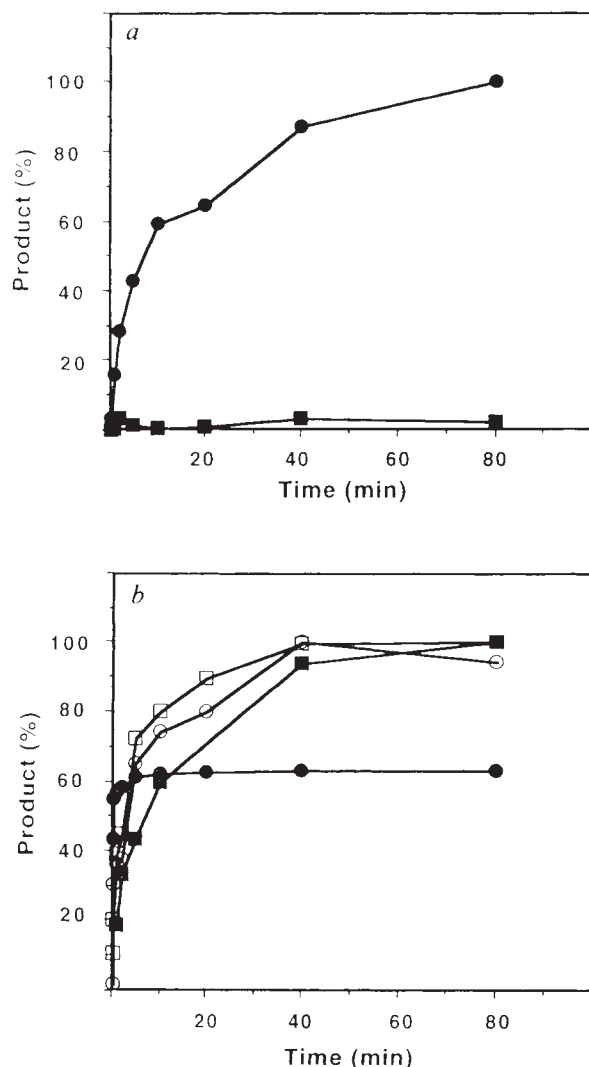


FIG. 4 Time course of *Eco*P15 restriction and modification of *a*, M13 DNA. ●, Restriction product; ■, modification product. *b*, M13 EM2 RF DNA; symbols as in *a*. Modification of M13 EM2 *Eco*P15-fragment A containing sites 1 and 3 (○), and fragment B containing site 2 (□).

METHODS 3.4 µg M13 or EM2 RF DNAs were linearized with *Bam*HI and subsequently labelled with [<sup>35</sup>S]dATP (1,000 Ci mmol<sup>-1</sup>) by standard procedures<sup>7</sup> using the Klenow fragment of DNA polymerase I. After phenol/chloroform extraction and precipitation with ethanol, the DNA was prewarmed in *Eco*P15 buffer<sup>8</sup> containing 1 mM ATP and 1 µM [methyl-<sup>3</sup>H]AdoMet (15 Ci mmol) in a total volume of 100 µl at 37 °C. The reaction was started by addition of 1 µg *Eco*P15. Aliquots of 10 µl were removed at the indicated time intervals. Reaction was stopped by addition of phenol and the DNA was separated from low-molecular-weight radioactive material by chromatography on Sephadex G-50. Ethanol precipitation was followed by electrophoresis in a 0.8% low-melting-temperature agarose gel. After cutting out the fragments, 3 volumes of 20 mM Tris-HCl, pH 8.0, 1 mM EDTA were added and incubated for 10 min at 70 °C. Finally the eluted fragments were counted in 2 ml scintillation cocktail (OptiPhase 'HiSafe' II, LKB, WALLAC). Restriction was determined as the ratio of cleaved to uncleaved fragments, measured by <sup>35</sup>S counting. Modification was analysed by the proportion of <sup>3</sup>H to <sup>35</sup>S counts.

§ To whom correspondence should be addressed.



TABLE 1 X-ray data collection and phasing statistics

| Data set             | Resolution (Å) | Number of observations | Unique reflections |                                | $R_{\text{sym}}$ (%) | $R_{\text{nat}}$ (%) |
|----------------------|----------------|------------------------|--------------------|--------------------------------|----------------------|----------------------|
|                      |                |                        | Number             | Percentage of possible         |                      |                      |
| Native               | 2.5            | 128,597                | 40,833             | 82.0 at 2.5 Å<br>95.2 at 3.5 Å | 6.5                  | —                    |
| EMTS                 | 3.5            | 44,757                 | 16,033             | 88.3                           | 5.4                  | 19.4                 |
| SIR statistics       |                |                        |                    |                                |                      |                      |
| Resolution range     |                |                        |                    |                                | 20–3.5 Å             |                      |
| Mean figure of merit |                |                        |                    |                                | 0.45                 |                      |
| Phasing power        |                |                        |                    |                                | 1.71                 |                      |
| Mean $\Delta\phi$    |                |                        |                    |                                | 70.9°                |                      |
| Averaging statistics |                |                        |                    |                                |                      |                      |
| Local symmetry       |                | 222 approximation      |                    | Refined pseudo 222             |                      |                      |
| Correlation          |                | 88.6%                  |                    | 94.9%                          |                      |                      |
| $R$ -factor          |                | 34.3%                  |                    | 18.3%                          |                      |                      |
| Mean $\Delta\phi$    |                | 52.6°                  |                    | 38.0°                          |                      |                      |

$R_{\text{sym}} (\sum |I_1 - I_2| / \sum I_2)$  and  $R_{\text{nat}} (\sum |F_1 - F_2| / \sum F_2)$  are the  $R$  factors between Friedel pairs within a data set and between the derivative and the native data sets. The single isomorphous replacement (SIR) mean figure of merit and phasing power are shown. Two sets of averaging statistics are given: those corresponding to the canonical 222 local symmetry and those corresponding to the correct local symmetry (see legend to Fig. 1). The correlation is calculated as  $\frac{1}{3} \sum_r \{ \sum_{i,j \neq i} \rho(T^i \cdot r) \rho(T^j \cdot r) \} / \{ \sum_i (\rho(T^i \cdot r))^2 \}$ , where  $\rho$  is the electron density,  $r$  runs through the points inside the envelope, and  $T^i$  is a local symmetry element ( $i=1, 4$ ). The  $R$  factor is calculated between  $F_{\text{obs}}$  and  $F_{\text{avg}}$ , where  $F_{\text{avg}}$  is the amplitude of the inverse Fourier transform of

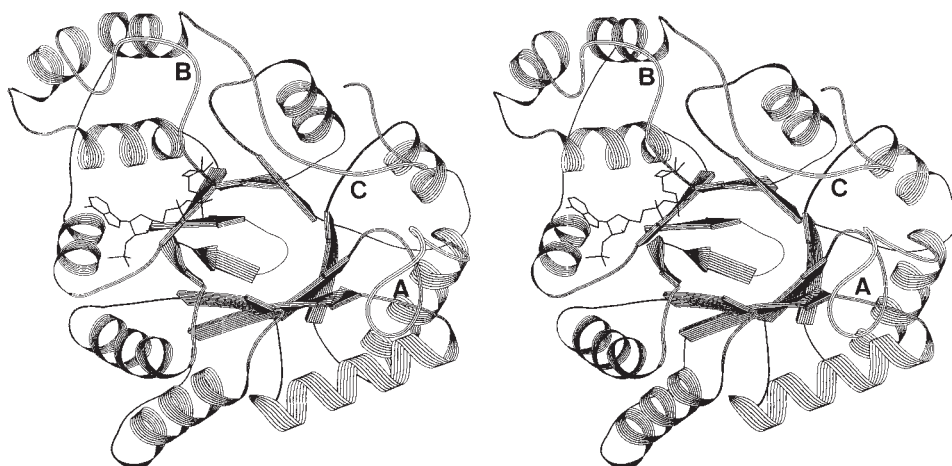
that it belongs to the expanding group of  $\alpha/\beta$  barrel enzymes and it is reasonable to assume that all the homologous proteins of this superfamily have a similar fold.

Pig lens aldose reductase is a single chain of 314 amino acids with an acetylated N-terminal serine. The enzyme is globular with a single structural domain, very similar to the archetypal structure of the triose phosphate isomerase  $\alpha/\beta$  barrel<sup>3</sup> (Fig. 1). When triclinic crystals of aldose reductase are soaked in a solution containing NADP(H), they crack and dissolve in a few minutes, suggesting that crystal packing is not compatible with

the averaged map. The phase error  $\Delta\phi$  is calculated between the current phase and the final model phase. Aldose reductase was purified from pig lens by homogenization of the lenses (25 mM bis-Tris, pH 6.2, 5 mM EDTA, 5 mM DTT (buffer A)) followed by centrifugation and fractionation of the supernatant using polyethylene glycol 3000 (brought to 20% (w/v) with PEG 3000). After centrifugation, the supernatant was loaded onto an anion exchanger (ZetaPrep cartridge QAE 100, equilibration: buffer A diluted 1.5 times in water, elution: 25 mM bis-Tris, pH 6.2, 50 mM NaCl, 1 mM EDTA, 5 mM DTT). Active fractions were pooled, concentrated by ultrafiltration (Amicon PM 10 membrane) and the buffer was changed to 25 mM piperazine chloride, pH 6.1, 2 mM DTT (buffer B) by several concentration/dilution cycles. The concentrate was then chromatofocused in the pH interval 5.5–4.5 (Mono-P HR 5/20 column, equilibrated in buffer B). Triclinic crystals of the apo-enzyme were grown at 4 °C by vapour diffusion in hanging drops as described<sup>20</sup>. The best crystals diffract beyond 2.0 Å resolution. Their unit cell parameters are  $a=81.3$  Å,  $b=85.9$  Å,  $c=56.6$  Å,  $\alpha=102.3^\circ$ ,  $\beta=103.3^\circ$ ,  $\gamma=79.0^\circ$ . Each cell contains four independent monomers arranged in a tetramer with pseudo 222 symmetry. A native data set to 2.5 Å resolution and a derivative data set to 3.5 Å resolution were collected from one single crystal each. Intensity data collection was at  $-5^\circ\text{C}$  with a Xentronics multiwire area detector equipped with a helium path and mounted on a Rigaku RU-200 rotating anode X-ray generator. Graphite monochromated radiation was used for data collection. The XDS program was used for raw data processing and reduction<sup>21</sup>. The structure has been solved by a combination of single isomorphous replacement, density modification and local symmetry averaging techniques (Fig. 2). Pig lens aldose reductase has been sequenced by a combination of chemical methods and mass spectrometry (A. Van Dorsselaer and M. Jaquinod, unpublished results) and about 90% of the sequence is now available. The sequence of the highly homologous human aldose reductase (about 88% identity) has been used for peptide alignment and structure determination whenever the porcine sequence was incomplete.

the binding of coenzyme. Attempts at cocrystallization yielded only thin needles that were not amenable to X-ray analysis. We therefore used ADPRP (2'-monophosphoadenosine-5'-diphosphoribose), a coenzyme analogue lacking the nicotinamide moiety of NADP(H), as a competitive inhibitor of pig lens aldose reductase with respect to NADPH binding (inhibition constant,  $K_i = 44$   $\mu\text{M}$ ). Thus this analogue can be used to localize the coenzyme binding site, with the caveat that the two compounds might bind differently and/or induce different local conformational changes. The binding site of ADPRP, a crevice

FIG. 1 Ribbon drawing<sup>26</sup> of the polypeptide backbone of aldose reductase showing the binding site of ADPRP. The aldose reductase structure comprises essentially an inner cylindrical core of 8 parallel  $\beta$ -strands surrounded by 8  $\alpha$ -helices. The chain proceeds around the barrel through the 8  $\beta\alpha$  units arranged in a sequential manner. At the N terminus two short antiparallel  $\beta$ -strands (residues 2–5 and 9–12) that are connected by a tight turn close off the bottom of the barrel. Three long loops, A, B and C, protrude from the top of the  $\alpha/\beta$  barrel at the C-terminal end of the  $\beta$ -strands, and partially cover a Y-shaped crevice. An extra piece of  $\alpha$ -helix is found between strand 7 and helix 7 (residues 230–239) and another after helix 8 (residues 281–288). The C-terminal loop C faces



loop A. The adenine-binding site is delineated by strands 7 and 8 and helices 7 and 8. The 2'-phosphate is directed towards the connection between strand 8 and helix 8. The pyrophosphate group points towards the centre of the barrel and interacts with sidechains of strands 7 and 8 and of the loop connecting strand 1 and helix 1. Although there are six cysteines in the sequence there are no disulphide bridges in this structure. Cys residues may be involved in the interconversion of two enzyme forms of aldose

reductase, which differ in their catalytic properties and their susceptibility to inhibitors<sup>27–30</sup>. Although the crystals used here were grown and stored for several weeks in the absence of a reducing agent, all cysteine residues are in the reduced state. Three cysteines are accessible to solvent and were modified by the organomercurial reagent ethylmercurithiosalicylate. They are located at the C-terminal end of strand 3 (C79) and in the C-terminal loop C (C297 and C302).

located in the C-terminal end of the barrel, has been found by difference Fourier analysis at 3.3 Å resolution of a crystal soaked with 25 mM ADPRP for 5 hours (Fig. 1). The electron density corresponding to the 'nicotinamide' ribose is very poorly defined in the difference Fourier map. But our results agree with the idea that NADPH binds in an open extended conformation, like dinucleotides binding to a Rossmann fold<sup>4</sup>. When ADPRP binds, loop B (Fig. 1) undergoes a limited conformational change towards the adenine moiety.

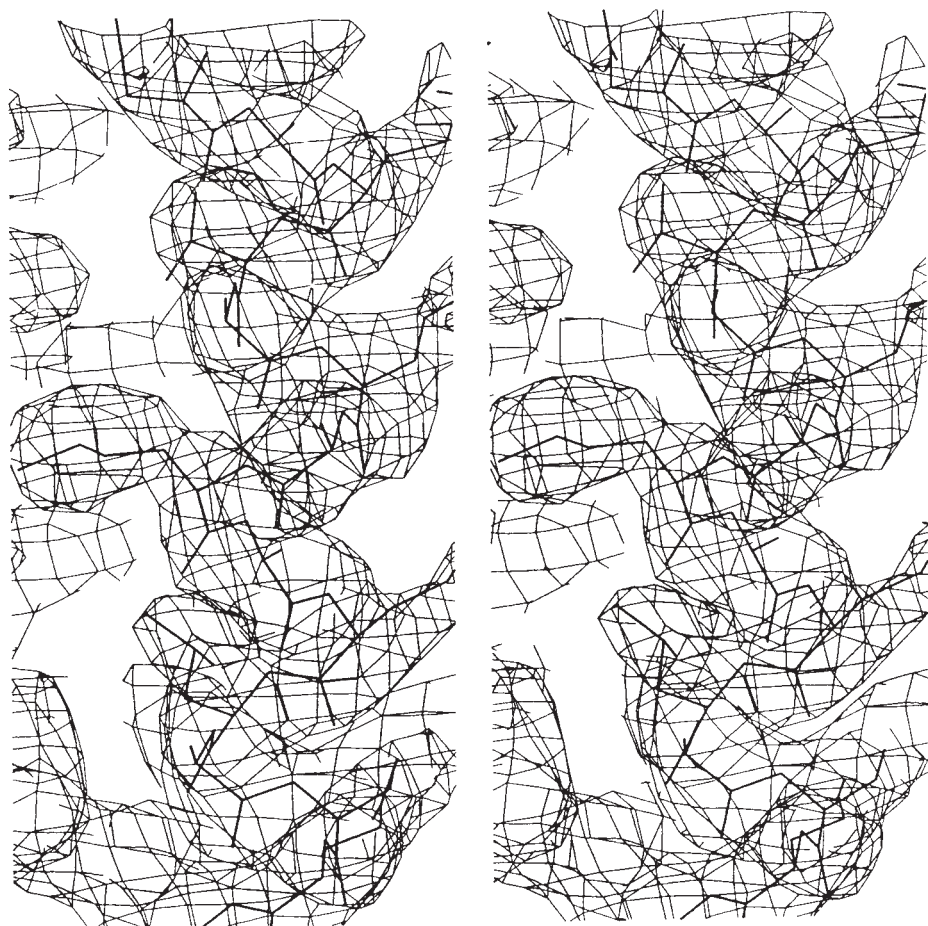
Other nucleotide-binding domains apart from the Rossmann fold have been noted, for example in the NADP-dependent enzymes dihydrofolate reductase<sup>15</sup> and *p*-hydroxybenzoate hydroxylase<sup>16</sup>. But the crystal structure of aldose reductase reveals that the  $\alpha/\beta$  barrel NADP-dependent oxidoreductases form a new structural family of NADPH-dependent enzymes. Conserved features with respect to other dehydrogenases are the localization of the binding site at the C-terminal end of parallel  $\beta$ -strands connected by  $\alpha$ -helices, and possibly the extended conformation of the bound coenzyme. Major structural differences result from the single domain structure of this monomeric enzyme and from the closed cylindrical  $\alpha/\beta$  barrel topology. In aldose reductase, residues of both the N-terminal

and the C-terminal regions are brought together to participate in the formation of the coenzyme binding site. Superposition of the coenzyme NAD bound to alcohol dehydrogenase and ADPRP bound to aldose reductase shows that the two binding domains approach the dinucleotides from opposite sides. Another striking difference with known nucleotide-binding domains lies in the position of the pyrophosphate moiety, which is generally located in the vicinity of the N terminus of at least one  $\alpha$ -helix. A favourable interaction between the negatively charged pyrophosphate group and the  $\alpha$ -helix dipole has been proposed to explain this binding<sup>17</sup>. We do not find any evidence for such an interaction here, stabilization being achieved through positively charged side chains.

The family of  $\alpha/\beta$ -barrel proteins shows a broad range of physiological and catalytic functions, substrate specificities and cofactor requirements<sup>18</sup>. Aldose reductase is the first example of an  $\alpha/\beta$  barrel enzyme using nicotinamide dinucleotide as cofactor. Like all  $\alpha/\beta$ -barrel enzymes examined so far, its active site is on the C-terminal end of the barrel and involves loop residues. It is interesting that the binding of ADPRP involves the same region of the  $\alpha/\beta$  barrel implicated in the binding of flavin mononucleotide to glycolate oxidase, flavocytochrome *b*<sub>2</sub>

FIG. 2 Stereo view of a typical zone of the final experimental map at 3.15 Å resolution. A portion of the final electron density after density averaging and phase extension from 3.50 to 3.15 Å is shown in the region of helix 4. The map is contoured at 1.0 standard deviation above the mean.

**METHODS.** Extensive search for heavy-atom derivatives yielded only one useful derivative, ethylmercurithiosalicylate. The heavy-atom structure (12 sites) was solved by a combination of manual inspection of the difference Patterson map, automatic search using the program HASSP<sup>22</sup> and location of additional sites in residual Fourier maps. Initial refinement of positions and occupancies was done with the program HEAVY<sup>23</sup>. At this stage, analysis of the heavy-atom constellation indicated that the arrangement of the sites could be described by a 222 point-group symmetry, confirming the results of the self-rotation function. The initial SIR phases were then improved using fourfold density averaging and solvent flattening, with the local program suite RMOL (developed by B. Rees). The resulting electron-density map at 3.5-Å resolution was sufficiently clear to allow the building of a partial model on a PS390 Evans and Sutherland graphic display with the program FRODO<sup>24</sup>. Sequence alignment was based on the identification of the N terminus and confirmed by the position of the three cysteine residues modified by ethylmercurithiosalicylate. One monomer was built in the averaged electron density and the non-crystallographic symmetry relationships were used to generate the whole tetramer. Rigid body refinement was then performed to refine the position and orientation of the four crystallographically independent monomers. It was followed by several energy minimization steps using the POWELL minimizer of the XPLOR package<sup>25</sup> with non-crystallographic symmetry restraints and data in ranges of increasing resolution (10.0–3.5, 10.0–3.0, 10.0–2.5 Å). At this stage the *R*-factor was 33.1%. After the rigid body refinement, the non-crystallographic symmetry showed a significant departure from a canonical 222 point-group symmetry. The new symmetry, together with improved molecular envelopes deduced from the current model,



was used for density averaging and phase extension from 3.5 to 2.90 Å. The entire model could be built at this stage. This model was refined by energy minimization followed by isotropic atomic B adjustment to a crystallographic *R*-factor of 23.2% for all data between 8.0- and 2.5-Å resolution with amplitudes greater or equal to 3 s.d. Although solvent molecules are clearly seen in the  $3F_o - 2F_c$  map, they have not been included yet in the refinement. The model has acceptable geometry with r.m.s. deviations from equilibrium of 0.024 Å for bond lengths, 4.6° for angles, 26.5° for dihedrals and 2.2° for improvers.



and trimethylamine dehydrogenase<sup>19</sup>.

Our results show that structural diversity prevails within the NADP-dependent enzyme family, even when function is closely related. On the other hand, it emphasizes the versatility of the  $\alpha/\beta$ -barrel scaffold, which often appears even in functionally unrelated proteins. Indeed, it is becoming evident that the number of stable folds used to achieve biological diversity is limited.

Preliminary clinical studies strongly support the value of aldose reductase inhibitors in the treatment of diabetic complications<sup>2</sup>. Knowledge of the three-dimensional structure of aldose reductase will enable more specific drugs to be designed so that therapy can eventually be improved. □

Received 7 August; accepted 6 November 1991.

1. Burg, M. B. *Kidney Int.* **33**, 635–641 (1988).
2. Dvornik, D. in *Aldose Reductase Inhibition. An Approach to the Prevention of Diabetic Complications* (McGraw-Hill, New York, 1987).
3. Banner, D. W. *et al. Nature* **255**, 609–614 (1975).
4. Rossmann, M. G., Moras, D. & Olsen, K. W. *Nature* **250**, 194–199 (1974).
5. Wermuth, B. *Prog. clin. biol. Res.* **174**, 209–230 (1985).

6. Bohren, K. M., Bullock, B., Wermuth, B. & Gabbay, K. H. *J. biol. Chem.* **264**, 9547–9551 (1989).
7. Chung, S. & LaMendola, J. *J. biol. Chem.* **264**, 14775–14777 (1989).
8. Nishimura, C., Wistow, G. & Carper, D. *Prog. clin. biol. Res.* **290**, 211–220 (1989).
9. Schade, S. Z. *et al. J. biol. Chem.* **265**, 3628–3635 (1990).
10. Watanabe, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 11–15 (1988).
11. Anderson, S. *et al. Science* **230**, 144–149 (1985).
12. Fujii, Y. *et al. J. biol. Chem.* **265**, 9914–9923 (1990).
13. Winters, C. J., Molowa, D. T. & Guzelian, P. S. *Biochemistry* **29**, 1080–1087 (1990).
14. Oechsner, U., Magdolen, V. & Bandlow, W. *FEBS Lett.* **238**, 123–128 (1988).
15. Wierenga, R. K., Drenth, J. & Schulz, G. E. *J. molec. Biol.* **167**, 725–739 (1983).
16. Bystroff, C., Oatley, S. J. & Kraut, J. *Biochemistry* **29**, 3263–3277 (1990).
17. Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. *Biochemistry* **24**, 1346–1357 (1985).
18. Farber, G. K. & Petsko, G. A. *Trends biol. Sci.* **15**, 228–234 (1990).
19. Lindqvist, Y., Brändén, C.-I., Mathews, F. S. & Lederer, F. *J. biol. Chem.* **266**, 3198–3207 (1991).
20. Rondeau, J. M. *et al. J. molec. Biol.* **195**, 945–948 (1987).
21. Kabsch, W. *J. appl. Crystallogr.* **21**, 916–924 (1988).
22. Terwilliger, T. C. & Kim, S.-H. *Acta crystallogr.* **A43**, 1–5 (1987).
23. Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39**, 813–817 (1983).
24. Jones, T. A. in *Computational Crystallography* 303–317 (Clarendon, Oxford, 1982).
25. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **25**, 458–460 (1987).
26. Priestle, J. *J. appl. Crystallogr.* **21**, 572–576 (1988).
27. Grimshaw, C. E. *et al. Biochemistry* **28**, 5343–5353 (1989).
28. Vander Jagt, D. L., Robinson, B., Taylor, K. K. & Hunsaker, L. A. *J. biol. Chem.* **265**, 20982–20987 (1990).
29. Del Corso, A. *et al. Arch. biochem. biophys.* **270**, 604–610 (1989).
30. Del Corso, A. *et al. Arch. biochem. biophys.* **283**, 512–518 (1990).

ACKNOWLEDGEMENTS. We thank A. Van Dorsselaer and M. Jaquinod for unpublished sequence data, B. Rees for his contribution to the phase problem and Y. S. Babu (of Biocryst) for discussions.

## Free $R$ value: a novel statistical quantity for assessing the accuracy of crystal structures

Axel T. Brünger

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, 06511, USA

THE determination of macromolecular structure by crystallography involves fitting atomic models to the observed diffraction data<sup>1</sup>. The traditional measure of the quality of this fit, and presumably the accuracy of the model, is the  $R$  value. Despite stereochemical restraints<sup>2</sup>, it is possible to overfit or 'misfit' the diffraction data: an incorrect model can be refined to fairly good  $R$  values as several recent examples have shown<sup>3</sup>. Here I propose a reliable and unbiased indicator of the accuracy of such models. By analogy with the cross-validation method<sup>4,5</sup> of testing statistical models I define a statistical quantity ( $R_T^{\text{free}}$ ) that measures the agreement between observed and computed structure factor amplitudes for a 'test' set of reflections that is omitted in the modelling and refinement process. As examples show, there is a high correlation between  $R_T^{\text{free}}$  and the accuracy of the atomic model phases. This is useful because experimental phase information is usually inaccurate, incomplete or unavailable. I expect that  $R_T^{\text{free}}$  will provide a measure of the information content of recently proposed models of thermal motion and disorder<sup>6–8</sup>, time-averaging<sup>9</sup> and bulk solvent<sup>10</sup>.

The most common measure for the quality of a crystal structure is the  $R$  value<sup>11</sup>,

$$R = \frac{\sum_{h,k,l} \|F_{\text{obs}}(h, k, l) - |k| F_{\text{calc}}(h, k, l)\|}{\sum_{h,k,l} |F_{\text{obs}}(h, k, l)|} \quad (1)$$

where  $h, k, l$  are the reciprocal lattice points of the crystal,  $|F_{\text{obs}}(h, k, l)|$  and  $|F_{\text{calc}}(h, k, l)|$  are the observed and calculated structure factor amplitudes, respectively.  $R$  is closely related to the crystallographic residual<sup>11</sup>

$$R' = \sum_{h,k,l} (|F_{\text{obs}}(h, k, l)| - |k| F_{\text{calc}}(h, k, l))^2 \quad (2)$$

which is a linear function of the negative logarithm of the likelihood of the atomic model assuming that all observations are independent and normally distributed<sup>12</sup>.  $R$  can be made arbitrarily small by increasing the number of model parameters

and subsequent refinement against  $R'$  (ref. 13), that is the diffraction data can be overfit without improvement or even worsening of the information content of the atomic model.

Crystallographic diffraction data are redundant to some degree, for example, refinement of the penicillopepsin crystal structure from *Penicillium janthinellum*<sup>14,15</sup> at 1.8 Å resolution with 50% of the diffraction data randomly omitted only results in a 0.3 Å root-mean-square (r.m.s.) difference to the atomic structure refined against the full data set (Fig. 1). In analogy to cross-validation<sup>4,5</sup> I thus propose to partition a unique set of the observed reflections into a 'test' set  $T$  and a 'working' set  $A$ , that is,  $T$  and  $A$  are disjoint and their conjunction is the full

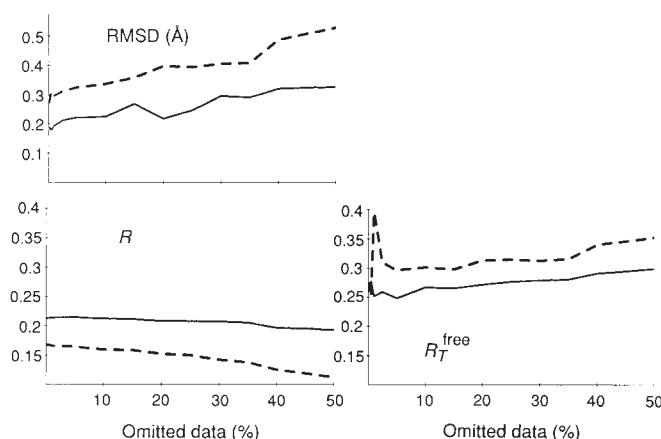


FIG. 1 SA-refinements of penicillopepsin<sup>14,15</sup> at 6–1.8 (solid lines) and 6–2.8 Å (dashed lines) resolution as a function of the percentage of omitted data.  $T$  was obtained by random selection from a unique set of all observed reflections.  $R$  is computed for the  $A$  set of reflections. The r.m.s. differences (RMSD) are computed between the structures refined against  $A$  and a unique set of all observed reflections. The RMSD is unequal to zero for 0% of the data omitted; this reflects the r.m.s. difference between two independently refined structures<sup>25</sup>. The penicillopepsin crystal structure<sup>14,15</sup> without water molecules and unit occupancy values was used as the starting point. Each refinement consisted of a slow-cooling protocol<sup>26</sup> using the program X-PLOR<sup>16,27</sup> starting at 1,000 K, overall B-factor refinement, and restrained individual B-factor refinement with the target values for the temperature factor deviations<sup>2</sup> of 1.5, 2, 2, 2.5 for bonded backbone, angle-related backbone, bonded sidechain, and angle-related sidechain atoms, respectively.

set of observed reflections. I refer to

$$R_T^{\text{free}} = \frac{\sum_{(h,k,l) \in T} \|F_{\text{obs}}(h, k, l) - k|F_{\text{calc}}(h, k, l)\|}{\sum_{(h,k,l) \in T} |F_{\text{obs}}(h, k, l)|} \quad (3)$$

as the free  $R$  value computed for the  $T$  set of reflections.  $T$  is omitted in the modelling process, for example in the case of crystallographic refinement<sup>2</sup> the residual to be minimized is given by

$$R'_A = \sum_{(h,k,l) \in A} (|F_{\text{obs}}(h, k, l) - k|F_{\text{calc}}(h, k, l)|)^2. \quad (4)$$

One would expect that  $R_T^{\text{free}}$  is less prone to overfitting than  $R$ . This concept can be applied to other statistical quantities, such as the standard linear correlation coefficient<sup>11</sup>. It can even be applied to crystal structures which have already been refined with all diffraction data included: refinement by simulated annealing (SA)<sup>16</sup> with  $T$  omitted will remove some of the memory towards  $T$ .

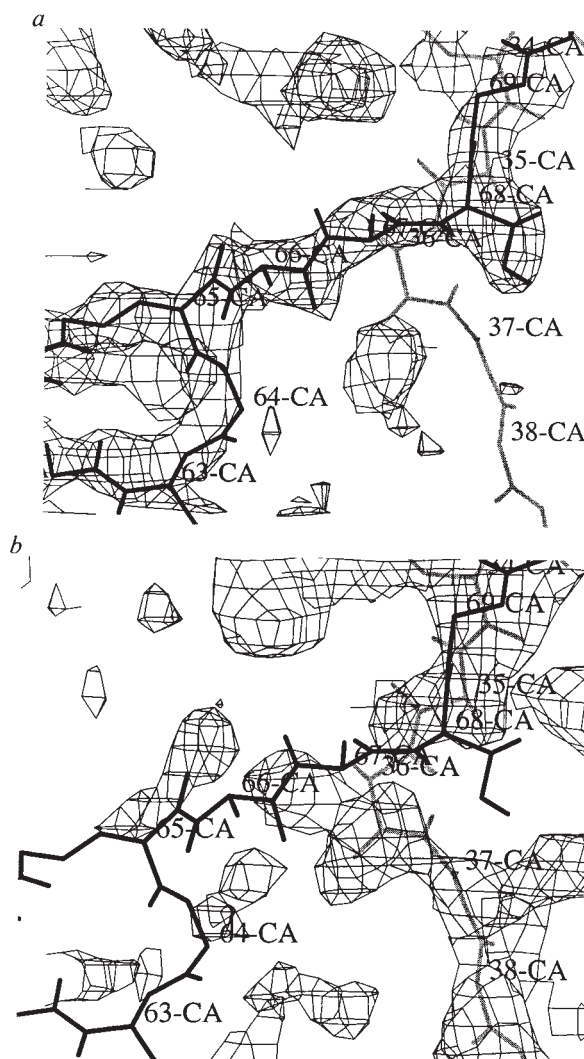
$R_T^{\text{free}}$  reflects the information content of the atomic model. Suppose both the atomic model and diffraction data are perfect, resulting in  $R = 0$ . Refinement against  $A$  as opposed to all data will not change the atomic model and thus  $R_T^{\text{free}} = 0$ . Suppose the data contain small errors and an atomic model is overfit to a very low  $R$  value by introducing a large number of free parameters. As the noise is independent among different reflections, overfitting against  $A$  will not bias  $R_T^{\text{free}}$ . A similar argument applies to the case of partially incomplete or incorrect atomic

models where the agreement with the diffraction data is improved by fitting noise.

The enhanced sensitivity of  $R_T^{\text{free}}$  with respect to model errors is illustrated in Fig. 2 which compares a portion of the correct<sup>17</sup> and incorrect<sup>18</sup> crystal structures of the plant ribulose-1,5-biphosphate carboxylase oxygenase (RuBisCO). Although the  $R$  difference between the correct and incorrect model is only 4% for comparable geometry, the  $R_T^{\text{free}}$  difference is 13%, suggesting that the incorrect model had been overfit. This is corroborated by the electron density maps in Fig. 2 which show a poorer agreement for the incorrect model. An alternative way to detect the errors in the RuBisCO crystal structure is provided by computing 'omit maps' with simulated annealing (A. Hodel, D. Eisenberg, S.-H. Kim and A.T.B., manuscript in preparation) which essentially is the real-space analogue to  $R_T^{\text{free}}$ .

Both  $R_T^{\text{free}}$  and the r.m.s. difference between the model refined against the complete data set and against  $A$  increase more or less monotonically as a function of the percentage of omitted data (Fig. 1). This is to be expected for terms that truly monitor the validity of a model.  $R$  decreases, which is paradoxical and misleading behaviour for an indicator of the models accuracy. As a compromise between avoiding fluctuations of  $R_T^{\text{free}}$  and maintaining small r.m.s. differences between refined models, I suggest  $T$  is obtained from a random selection of 10% of the observed reflections. The definition of  $R_T^{\text{free}}$  implies  $R_T^{\text{free}} > R$ ; the difference between  $R_T^{\text{free}}$  and  $R$  is uniformly distributed as a function of resolution (not shown).

FIG. 2 The region around residue 66 of the small subunit of RuBisCO. The correct<sup>17</sup> structure is shown in black, whereas the incorrect<sup>18</sup> structure, which involved the nearly backwards tracing of the polypeptide chain of the small subunit, is shown in grey. Superimposed are  $\sigma_A$ -weighted<sup>28</sup>  $2F_o - F_c$  electron density maps with phases computed from the correct model (a,  $R = 0.16$ ,  $R_T^{\text{free}} = 0.34$ ) and incorrect model (b,  $R = 0.2$ ,  $R_T^{\text{free}} = 0.47$ ) shown at 2.5 Å resolution for a contour level of  $1\sigma$ . The maps are ordinary omit maps (for a review of omit map techniques, see A. Hodel, D. Eisenberg, S.-H. Kim and A.T.B., manuscript in preparation), that is residues 36–47 and all residues within 5 Å of this loop were removed in the phase calculation.  $T$  was obtained by a 10% random selection from the observed reflections. SA-refinements and restrained B-factor refinements were done at 2.5 Å resolution using A. The r.m.s. deviations of bond lengths and bond angles from ideal were 0.02 Å and 4°, respectively, for the correct structure whereas they were 0.03 Å and 5° for the incorrect structure.





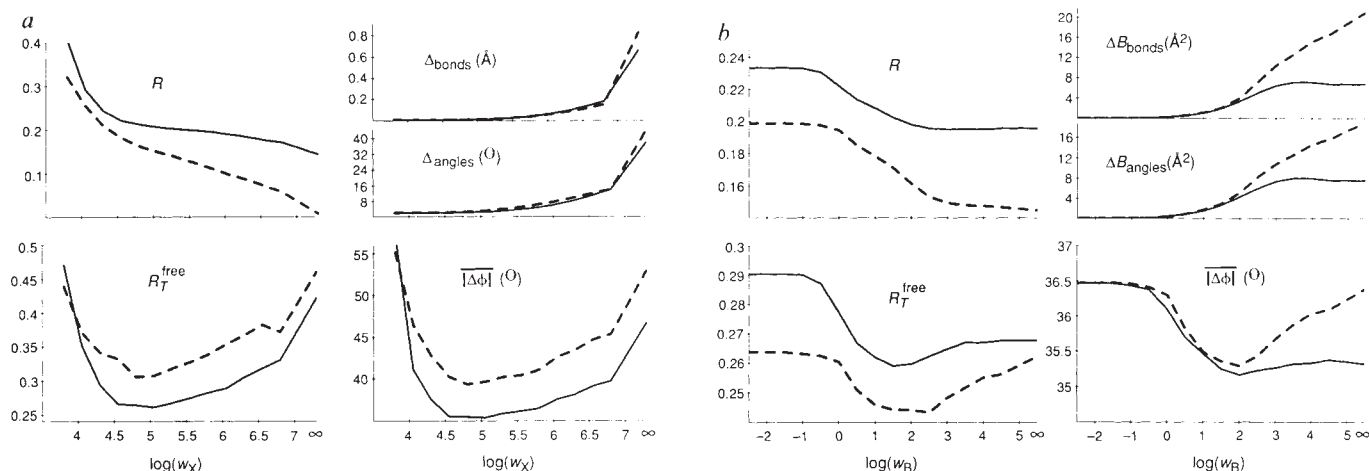


FIG. 3 *a*, SA-refinements of penicillopepsin<sup>14,15</sup> at 6–1.8 (solid lines) and 6–2.8 Å (dashed lines) resolution as a function of  $w_x$  (equation (5)) with  $R'$  replaced by  $R'_A$  (equation (4)).  $T$  was obtained by a 10% random selection.  $R$  was computed for  $A$ .  $\Delta_{\text{bonds}}$  and  $\Delta_{\text{angles}}$  are the r.m.s. deviations of bond lengths and bond angles from their ideal values.  $|\Delta\Phi|$  is the figure-of-merit weighted mean phase difference between model phases and the most probable MIR phases at 6–2.8 Å resolution. Details of the penicillopepsin model and refinement procedure are the same as in Fig. 1. The standard

linear correlation coefficient between the  $R_T^{\text{free}}$  and  $|\Delta\Phi|$  graphs is 0.98 for both resolution ranges. *b*, Restrained B-factor refinements of penicillopepsin as a function of  $w_B$  (equation (6)) with  $R'$  replaced by  $R'_A$  (equation (4)).  $\Delta B_{\text{bonds}}$  and  $\Delta B_{\text{angles}}$  are the r.m.s. deviations between B-factors of atoms sharing a covalent bond or bond angle, respectively.  $w_B = \infty$  represents the completely unrestrained case whereas  $w_B \rightarrow -\infty$  represents refinement of a single overall B-factor.

Diffraction data and prior knowledge are often combined as is the case in restrained least-squares refinement of atomic positions<sup>2</sup> that can be viewed as minimization of the atomic coordinates against a cost function<sup>19,20</sup>

$$C = w_x R' + E_{\text{chemical}} \quad (5)$$

where  $w_x$  is a weight and  $E_{\text{chemical}}$  is a geometric<sup>19</sup> or empirical<sup>21</sup> energy function which has been made unitless by multiplication of a conversion factor. If  $w_x$  chosen is too small, too much emphasis is put on the geometry as provided in  $E_{\text{chemical}}$ , which results in an inaccurate  $R$  value. If  $w_x$  is chosen too large, the structure will be overfit to a very good  $R$  value, but the geometry of the structure becomes severely distorted. The optimal choice of  $w_x$  cannot be obtained by linear hypothesis tests<sup>13</sup> because of the presence of nonlinear restraints, such as repulsive contact functions<sup>2</sup>.  $R_T^{\text{free}}$  is not subject to such limitations.

A series of positional refinements of the penicillopepsin structure<sup>14,15</sup> produced a minimum for  $R_T^{\text{free}}$  at  $\log(w_x) = 5$ , independent of the resolution range used (Fig. 3*a*). At this minimal value the r.m.s. deviation of bond lengths and bond angles from ideality are 0.013 Å and 2.5°, respectively. As an independent determination of the optimal  $w_x$  I used the multiple isomorphous replacement (MIR) phases at 6–2.8 Å resolution<sup>14,15</sup>; these phases were of exceptional quality with a figure of merit of 0.9. Experimental phase information is normally less accurate, incomplete or missing.  $R_T^{\text{free}}$  is highly correlated with the mean difference between the model and the MIR phases ( $|\Delta\Phi|$ ) (Fig. 3*a*).  $R_T^{\text{free}}$  thus yields the optimal choice for  $w_x$  without reference

to experimental phase information or to expected deviations of the geometry from ideality. The resulting relatively tight geometry is a consequence of the diffraction data, not of the geometric or empirical energy function.

Restrained temperature factor refinement<sup>22</sup> poses a similar

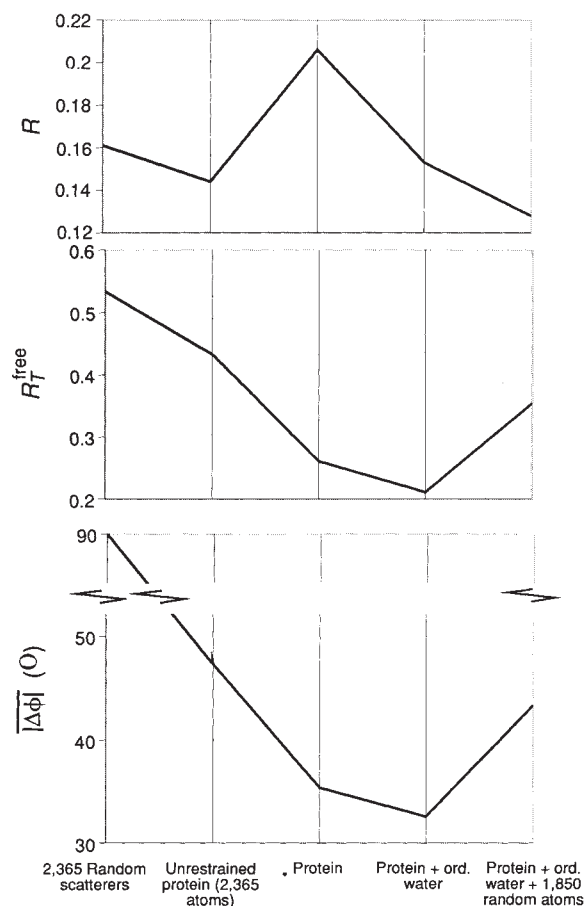


FIG. 4 '2,365 Random scatterers' consists of 2,365 oxygen atoms with a reduced van der Waals' radius of 1.57 Å randomly placed in the asymmetric unit of the crystal. 'Unrestrained protein', consists of the same scatterers placed near the non-hydrogen positions of the protein portion of the penicillopepsin structure refined with chemical restraints (equation (5)). 'Protein + ord. water' includes an additional 314 ordered water molecules. 'Protein + ord. water + 1,850 random atoms' includes an additional 1,850 oxygen atoms randomly placed in the bulk solvent region. The definition of  $T$  and  $|\Delta\Phi|$  is identical to Fig. 3. Each refinement consisted of two iterations of SA-refinement and restrained B-factor refinement as detailed in Fig. 1. The standard linear correlation coefficient between the  $R_T^{\text{free}}$  and  $|\Delta\Phi|$  graphs is 0.98.

problem as positional refinement. It consists of minimization of a cost function

$$C = w_B R' + \sum_{(i,j) \text{ bonds}} \frac{(B_i - B_j)^2}{\sigma_{\text{bonds}}^2} + \sum_{(i,j,k) \text{ angles}} \frac{(B_i - B_k)^2}{\sigma_{\text{angles}}^2} \quad (6)$$

where  $B_i$  is the temperature factor of atom  $i$  and the summations are carried out over all covalent bonds and bond angles<sup>22</sup>.  $R_T^{\text{free}}$  is highly correlated with  $|\Delta\Phi|$  and thus determines the optimal choice of  $w_B$  (Fig. 3b).

The information content of a random distribution of scatterers is obviously minimal, although it can be refined to a very low  $R$  value (Fig. 4);  $R_T^{\text{free}}$  stays at 54% which is close to the random limit of 59% for an acentric space group<sup>11</sup>. Unrestrained refinement with a model consisting of the same scattering starting at the positions of the non-hydrogen protein atoms yields  $R_T^{\text{free}} = 43\%$  (Fig. 4). Thus,  $R_T^{\text{free}}$  can distinguish between a distribution of scatterers that is close to the crystal structure and a random distribution, both of which can be refined to a very low  $R$ . Inclusion of chemical restraints increases  $R$  somewhat while greatly decreasing both  $R_T^{\text{free}}$  and  $|\Delta\Phi|$ , thus improving the information content of the model (Fig. 4). Inclusion of ordered water molecules lowers  $R$ ,  $R_T^{\text{free}}$  and  $|\Delta\Phi|$  (Fig. 4). Refinement of randomly placed scatterers in the bulk solvent region of the crystal lowers  $R$  while increasing both  $R_T^{\text{free}}$  and  $|\Delta\Phi|$ , thus decreasing the information content.

$R_T^{\text{free}}$  represents a reliable and unbiased parameter by which to evaluate the information content of a model produced by X-ray crystallography. It is not restricted to high-resolution diffraction data: tests carried out both at 6–2.8 Å and at 6–1.8 Å resolution produce large correlations between  $R_T^{\text{free}}$  and  $|\Delta\Phi|$  (Fig. 3). The observation that  $R_T^{\text{free}}$  can distinguish between a random distribution of scatterers and distribution close to the protein suggests applications to *ab initio* phasing. The increase of  $R_T^{\text{free}}$  on modelling the bulk solvent region of the penicillopepsin structure with stationary atoms confirms the disordered character of bulk solvent. A similar approach might be useful for the three-dimensional structure determined by solution NMR<sup>23,24,29</sup> if sufficient redundancy in the data and accuracy of the NOE (nuclear Overhauser effect) intensities can be achieved. □

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

#### Categories of paper

**Review articles** survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

**Articles** are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*. Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarizes its results and implications. Articles have fewer than 50 references and may contain a few subheadings of two or three words.

**Letters** are short reports of outstanding novel findings, whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

**Commentary articles** deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

**News and Views articles** inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the

News and Views editor.

**Scientific Correspondence** is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

#### Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required; for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

**Titles** say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

**Abstracts** should be marked individually and clearly with the author's name and, where known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 40 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome.

**Colour artwork.** A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

**Figure legends** should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

**References** are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included: reference to books should include publisher, place and date.

**Abbreviations, symbols, units and Greek letters** should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

**Acknowledgements** are brief; grant and contribution numbers are not allowed.

**Submission.** Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and value \$5 to prevent the imposition of import duty and value-added tax (VAT).

Received 14 August; accepted 30 October 1991.

- Jensen, L. H. *Meth. Enzym.* **115**, 227–234 (1985).
- Hendrickson, W. A. *Meth. Enzym.* **115**, 252–270 (1985).
- Bränden, C. I. & Jones, A. *Nature* **343**, 687–689 (1990).
- Mosteller, F. & Tukey, J. W. *Data Analysis and Regression: a Second Course in Statistics* (Addison-Wesley, Reading, Massachusetts, 1977).
- Efron, B. & Tibshirani, R. *Science* **253**, 390–395 (1991).
- Diamond, R. *Acta crystallogr.* **A46**, 425–435 (1990).
- Kuriyan, J. & Weiss, W. I. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2773–2777 (1991).
- Kuriyan, J. *et al.* *Proteins* **10**, 340–358 (1991).
- Gros, P., Gunsteren, W. F. & Hol, W. G. *J. Science* **249**, 1149–1152 (1990).
- Badger, J. & Caspar, D. L. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 622–626 (1991).
- Stout, G. H. & Jensen, L. H. in *X-ray Structure Determination, a Practical Guide* 2nd edn 343–378 (Wiley, New York, 1989).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. in *Numerical Recipes* 498–504 (Cambridge University Press, 1986).
- Hamilton, W. C. *Acta crystallogr.* **18**, 502–510 (1965).
- Hsu, I.-N., Delbare, L. T. J., James, M. N. G. & Hofmann, T. *Nature* **266**, 140–145 (1977).
- James, M. N. G. & Sielecki, A. R. *J. molec. Biol.* **163**, 299–361 (1983).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Curni, P. A. M., Schreuder, H., Cascio, D., Sweet, R. M. & Eisenberg, D. *J. biol. Chem.* (in the press).
- Chapman, M. S. *et al.* *Science* **241**, 71–74 (1988).
- Brünger, A. T. *A. Rev. Phys. Chem.* **42**, 197–233 (1991).
- Jack, A. & Levitt, M. *Acta crystallogr.* **A34**, 931–935 (1978).
- Brooks, B. R. *et al.* *J. comp. Chem.* **4**, 187–217 (1983).
- Konnert, J. H. & Hendrickson, W. A. *Acta crystallogr.* **A36**, 344–349 (1980).
- Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon, Oxford, 1987).
- Wüthrich, K., *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
- Brünger, A. T. *J. molec. Biol.* **203**, 803–816 (1988).
- Brünger, A. T., Krukowski, A. & Erickson, J. *Acta crystallogr.* **A46**, 585–593 (1990).
- Brünger, A. T. *X-PLOR*, Version 2.1, Yale University (1990).
- Read, R. *Acta crystallogr.* **A42**, 140–149 (1986).
- Clore, G. M. & Gronenborn, A. M. *Science* **252**, 1390–1399 (1991).

ACKNOWLEDGEMENTS. I thank M. N. G. James and A. R. Sielecki for providing the diffraction data and coordinates of the penicillopepsin structure, D. Eisenberg for providing the data and coordinates of RuBisCO and the Pittsburgh Supercomputer Center for support.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse.

I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom  
☐ OnType ☐ OCLC ILL Subsystem

☐ Other (please specify) \_\_\_\_\_

☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name \_\_\_\_\_

Title \_\_\_\_\_

Institution/Company \_\_\_\_\_

Department \_\_\_\_\_

Address \_\_\_\_\_

City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_

Phone (\_\_\_\_\_) \_\_\_\_\_

Mail to: University Microfilms International  
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106



# Antigenized antibodies

Maurizio Zanetti

A new process, antigenization of antibodies, consisting of the expression of oligopeptides in the hypervariable loops of an antibody molecule is described. The potential applications of antigenized antibodies are discussed.

CONSIDERABLE progress has been made in understanding the molecular mechanisms and the cellular events involved in the antigenicity of proteins as well as the requirements for the successful induction of humoral and cellular responses ('immunogenicity'). Whereas new methods have been developed for the synthesis of protein antigens, their practical utilization is still hampered by a poor understanding of the relation between amino acid sequence, protein folding and function at the level of the appropriate cell receptor(s). Over the past 10 years, emphasis has been placed on synthetic peptides as a convenient, man-made replica of short amino acid sequences of known antigen structures<sup>1-3</sup>. In spite of the initial enthusiasm, shortcomings to the synthetic peptide approach have slowly become apparent. Among these are poor predictability of the tertiary structure<sup>4</sup> and scarce immunogenicity<sup>5</sup>. Peptides in solution likely exist in several metastable conformations, which may or may not approximate that necessary for optimal receptor binding (B-cell receptor, and possibly T-cell receptor and major histocompatibility gene products) in case fidelity at the three-dimensional level is required. Thus, oligopeptide epitopes for immunological use need not only to reproduce the primary structure, but should also adhere to some intrinsic conformational requirements.

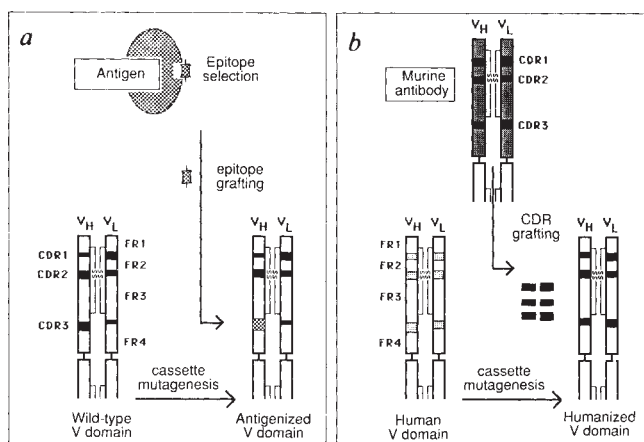
Protein engineering techniques<sup>6</sup> have made it possible to express oligopeptide epitopes of biological interest in a variety of systems. These include bacterial flagellar filaments<sup>7,8</sup> and fimbrial subunits<sup>9</sup>, and hepatitis B surface antigen<sup>10</sup>. Some attempts have already been made to test the ability of such recombinant proteins to serve as antigens for the elicitation of immunity directed at the oligopeptide epitope<sup>7</sup>. More recently, peptides have been expressed in the PIII protein on the surface of filaments coliphage<sup>11-14</sup>. Whereas this approach enables one to generate large numbers of sequences among which to select those reactive with a given ligand, it may be of limited use as a source of antigens needed to address a variety of biological questions.

## Antibody antigenization

Complementarity-determining region (CDR) loops of an antibody molecule are a new possibility for expressing oligopep-

FIG. 1 Schematic view of the process for the generation of antigenized antibodies (a) and, by comparison, of humanized antibodies (b). The representation underlines the fundamental difference between the two processes in that antigenization extends to epitopes of any protein antigens, whereas humanization refers to grafting the CDRs of one antibody, for example, a murine monoclonal

antibody, into the framework of human immunoglobulin. CDR, complementarity-determining regions; FR, framework region; V<sub>H</sub>, variable domain of the heavy chain; V<sub>L</sub>, variable domain of the light chain; and V, variable (domain).



tide epitopes in an immunologically-accessible and conformationally-suitable way. The new process, 'antibody antigenization', consists of grafting peptide epitopes derived from antigens other than immunoglobulins into antibody CDR loops<sup>15</sup>. The rationale is provided by physico-chemical considerations on antibody structure and what is known on the antigenicity of immunoglobulins.

The structure of the variable (V) domains of immunoglobulin is well characterized<sup>16</sup>. Framework regions (FR) organized as  $\beta$ -strands are interconnected by, and support, CDR loops; three for the V region of the heavy (H) and three for the V region of the light (L) chain. Whereas FR are relatively conserved among different antibodies, CDR loops at the top of the V domains vary in sequence and length<sup>17</sup>, and are to some extent independent from the physico-chemical constraints that maintain the packing of V<sub>H</sub> and V<sub>L</sub>  $\beta$ -sheets in a conserved framework<sup>18</sup>. The CDR variability accounts for most of the antigen-binding properties of antibodies. Most antigenic determinants of antibody V regions map to CDR loops, although residues in the FR may also participate. Antigenically-active loops of V domains are exposed areas at the surface of the molecule and both convexity and relative rigidity are important factors<sup>19</sup>. The conformation of the six CDR loops is, therefore, the principal determinant of the architecture of the antigen-binding site, and the major contributor to antibody antigenicity.

Within a model where the  $\beta$ -sheets that constitute the FR serve as a scaffold for CDR loops of various size and composition, it becomes apparent that CDR loops could be an ideal site for the integration and expression of oligopeptide epitopes of proteins other than immunoglobulins — antibody antigenization (see Fig. 1a). This process differs conceptually from that based on swapping CDR loops between two antibody molecules to transfer the antigen-binding of a murine or rat antibody to a human framework<sup>20-22</sup> in order to minimize the immune response to heterologous determinants — humanized antibodies<sup>23,24</sup> (see Fig. 1b).

## Testing the hypothesis

The antigenicity and immunogenicity of short amino acid sequences of exogenous antigens grafted into antibody CDR loops has been successfully tested in several systems. In the first instance, the third CDR of the heavy (H) chain of a murine V<sub>H</sub> was engineered to encompass the proline-rich tetrapeptide repeat (NANP)<sub>3</sub>, a B-cell epitope of *Plasmodium falciparum* malaria circumsporozoite protein<sup>25</sup>. The surface conformation of the V<sub>H</sub> domain of the antigenized antibody was probed with a monoclonal antibody specific for the (NANP)<sub>3</sub> epitope in the parasite antigen. We deduced that the structure assumed by the (NANP)<sub>3</sub> peptide within the V<sub>H</sub> domain was immunologically similar to the native antigen and, comparatively, more reactive than a synthetic peptide of identical

amino acid composition. The antigenized antibody was also used to immunize rabbits and mice and proved capable of eliciting a humoral response to (NANP), synthetic peptide and *P. falciparum* parasite in all animals immunized<sup>26</sup>. The induced antibodies efficiently inhibited the invasion of cultured liver cells by *P. falciparum* parasite *in vitro*. The effect can be interpreted within the framework of idiotype-anti-idiotype interactions with the antigenized antibody functioning as an 'internal image' mimicking antigen function by virtue of shared amino acid sequence<sup>27-29</sup>.

In another system, an antibody was constructed with a CDR loop encompassing ten amino acid residues from the sequence of a neuronal *c-src* protein. Rabbit antibodies generated against the antigenized antibody reacted with the neuronal *c-src* protein encoded by a baculovirus vector, specifically stained neuronal cells overexpressing the *c-src* protein, and immunoprecipitated a neuronal *c-src* protein active in a kinase assay (P. Rigaudy *et al.*, manuscript in preparation).

In a third system currently under study, an antibody antigenized to express a discrete portion of the human CD4 receptor proved to be a powerful immunogen for the induction of antibodies that react with native CD4 on T lymphocytes and block HIV-mediated syncytia formation *in vitro* (P. Lanza *et al.*, manuscript in preparation). It appears, therefore, that CDR loops are a good site for expressing oligopeptides in an immunologically accessible and functional way. The implication is that peptides immobilized within the immunoglobulin fold acquire conformational stability by assuming only a few of the several possible conformations allowed in solution.

## Looking ahead

Using CDR loops of both the H and L chains one can construct molecules with more than one epitope from the same antigen, for example, two epitopes from distinct parts of the same pathogen (like the outer coat and core protein of a virus), or from two different stages of the life cycle of the same pathogen (as could be the case for the sporozoite and merozoite stages of the malaria parasite). Using programs for three-dimensional modelling, it will be possible to guide the design and construction of binding domains for ligands, receptors or enzymes requiring more than one loop to carry a biological function.

Apart from immunogenicity at the B-cell level, antigenized antibodies may be exploited for presenting epitopes integrated in the CDR loops to T cells. Initial studies indicate that antibodies antigenized with selected epitopes of

microbial pathogens can induce epitope-specific T-helper and T-cytotoxic cells. Therefore, this new strategy allows one to construct immunogens tailored to a particular arm of the immune system. An advantage over conventional, recombinant and synthetic peptide is the possibility of targeting monocytes and macrophages via the Fc portion of the molecule. Antigenized antibodies carrying mutant structures of a reference epitope can be invaluable reagents in evaluating the reactivity of antibodies from protected and non-protected individuals, a step that is crucial in assessing the chemical characteristics of epitopes necessary for protection<sup>30</sup>. By X-ray diffraction analysis, antigenized Fab regions can facilitate resolving the three-dimensional structure of selected epitopes. Applied to receptor, ligand or enzyme substrate interaction, antigenized antibodies can serve to identify which among various candidate loops or peptidic sequence directly mediate binding or catalysis. The large number of potential applications for antigenized antibodies contrasts the relative simplicity involved in both the engineering and purification steps, that is, the chemical structure of the CDR loops can be easily programmed *in vitro* and the purification relies on conventional methods.

In conclusion, the experiments discussed above validate the new principle that determinants of proteins other than immunoglobulins can be made part of the structure of antibody molecules and be recognized as antigenic determinants by the appropriate receptor on cells of the immune system. In the years to come, the range of applications for antigenized antibodies may well involve new approaches to human therapy and prophylaxis as well as alternative strategies to understanding basic, structural aspects of protein-protein interactions. □

Maurizio Zanetti is in the Department of Medicine and Cancer Center, University of California — San Diego, San Diego, California 92103, USA. For more information, fill in reader service number 100.

- Atassi, M. Z. *Molec. cell. Biochem.* **32**, 21–43 (1980).
- Lerner, R. A. *Scient. Am.* **248**, 66–74 (1983).
- Geysen, H. M. *Immunol. Today* **6**, 364–369 (1985).
- Spangler, B. D. *J. Immunol.* **146**, 1591–1595 (1991).
- Newmark, P. *Nature* **305**, 9 (1983).
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367–382 (1987).
- Newton, S. M., Jacob, C. O. & Stocker, B. A. *Science* **244**, 70–72 (1989).
- Kuwajima, G. *et al. Bio/Technology* **6**, 1080–1083 (1988).
- Jennings, P. A., Bills, M. M., Irving, D. O. & Mattick, J. S. *Protein Eng.* **2**, 365–369 (1989).
- Rutgers, T. *et al. Bio/Technology* **6**, 1065–1070 (1988).
- Reidhaarlson, J. F. & Sauer, R. T. *Science* **241**, 53–57 (1988).
- Cwirila, S. E., Peters, E. A., Barrett, R. W. & Dower, W. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6378–6382 (1990).
- Devlin, J. J., Panganiban, L. C. & Devlin, P. E. *Science* **249**, 404–406 (1990).

- Scott, J. K. & Smith, G. P. *Science* **249**, 386–390 (1990).
- Zanetti, M., Billella, R. & Sollazzo, M. in *Anti-Idiotypic Vaccines* (ed. Cazenave, P. A.) 123–134 (Springer, New York, 1991).
- Alzari, P. M., Lacombe, M. – B. & Poljak, R. J. *A. Rev. Immunol.* **6**, 555–580 (1988).
- Kabat, E. A., Wu, E. T., Reidmiller, M., Perry, H. M. & Gottesman, K. S. *Proteins of Immunological Interest* (U.S. Department of Health and Human Services, Bethesda, 1987).
- Chothia, C., Novotny, J., Brucoleri, R. & Karplus, M. *J. molec. Biol.* **186**, 651–663 (1985).
- Novotny, J. & Franek, F. *Nature* **258**, 641–643 (1975).
- Verhoeven, M., Milstein, C. & Winter, G. *Science* **239**, 1534–1536 (1988).
- Jones, T. P., Dear, P. H., Foote, J., Neuberger, M. & Winter, G. *Nature* **321**, 522–525 (1986).
- Riechmann, L., Clark, M., Waldmann, H. & Winter, G. *Nature* **332**, 323–327 (1988).
- Tempest, P. R., *et al. Bio/Technology* **9**, 266–271 (1991).
- Co, M. S., Deschamps, M., Whitley, R. J. & Queen, C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2869–2873 (1991).
- Sollazzo, M., Billella, R. & Zanetti, M. *Protein Eng.* **4**, 215–220 (1990).
- Billella, R., Hollingdale, R. M. & Zanetti, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4713–4717 (1991).
- Ollier, P. R. S., Somme, G., Theze, J. & Fougereau, M. *EMBO J.* **4**, 3681–3688 (1985).
- Bruck, C., *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 6578–6582 (1986).
- Van, C. V., Naeve, C. W. & Metzger, D. W. *J. exp. Med.* **167**, 1841–1848 (1988).
- Hoffman, S., Nussenzweig, V., Sadoff, J. C. & Nussenzweig, R. S. *Science* **252**, 520–521 (1991).

## ADVERTISEMENTS

### A new label....



### ....unique products

Whether it's 25µL of a monoclonal antibody, 50pmol of an oligonucleotide probe or 5mg of a neurochemical.

Now they've no place to hide!

AFFINITY Research Products Ltd., 10 Goswell Industrial Estate, Ilkeston, DE7 5UG, UK

Tel: (+44/0)602 442232; Fax: (+44/0)602 442313

Specialising in Neuroscience Research

Reader Service No. 17

## BIANORM EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.



Reader Service No. 49



# Immunochemicals to order

A new chemiluminescent substrate for alkaline phosphatase, a custom monoclonal anti-peptide antibody service and a host of monoclonal antibodies — research tools for the immunologist.

CSPD ( $C_{18}H_{20}ClO_2PNa_2$ ) is the new high-performance dioxetane **chemiluminescent substrate for alkaline phosphatase** from Tropix (*Reader Service No. 101*). The substrate should be of interest to molecular biologists who use Southern blotting and DNA sequencing techniques and protein chemists who use ELISA and western blotting procedures. CSPD offers several advantages over available chemiluminescent substrates. First, its rapid light emission kinetics permits greater throughput for solution-based clinical assays and research methods, Tropix says. CSPD also enables researchers to obtain quicker test results on membrane surfaces due to its faster signal half-life. Second, CSPD offers enhanced detection levels of enzyme conjugates as compared to the company's AMPPD substrate. This, Tropix says, yields higher sensitivity in both clinical diagnostic assays and research applications. Additional features of CSPD include very low nonspecific background in solution and on membrane surfaces, and excellent thermal stability in the presence of polymeric enhancers. Ultrapure CSPD is available in the following concentrations (10 mg ml<sup>-1</sup>): 0.5, 2.5, 5, 10, 25 and 100 ml.

The new AS/AP Plus ultrasensitive **immunostaining kit** from Accurate Chemical and Scientific is designed for applications where the antibody exhibits a weak affinity for its epitope, or is difficult or expensive to acquire (*Reader*



STANDARD AS/AP  
Dilution 1:200

AS/AP PLUS  
Dilution 1:1,000

Ultrasensitive immunostaining kit.

*Service No. 102*). Use of the AS/AP Plus kit should yield better results and save money on expensive antibodies by allowing a higher dilution to be used, Accurate says. The AS/AP Plus kit is also available as a polyvalent kit that detects both mouse monoclonal and rabbit polyclonal primary antibodies.

ImmunoQuest ELISA assay kits from Blake Biotechnologies are designed to provide a quick and accurate **measurement of human secretory and serum IgA** (*Reader Service No. 103*). The kits

measure IgA in the low-nanogram range with a dynamic range of four orders of magnitude, the company says. The horseradish peroxidase-based kits contain Protein B, which binds to the Fc region of human IgA, with no detectable binding to other human immunoglobulins or serum proteins. ImmunoQuest kits, with convenient strip wells, are easy to use and make a useful test for IgA deficiency. Assay kits for antigen-specific IgA are under development. Blake Biotechnologies also sells conjugated and recombinant Protein B, and monoclonal antibodies to human IgA.

## Services and shelf stockers

Multiple Peptide Systems has recently expanded its range of immunological services with the introduction of a new proprietary **monoclonal anti-peptide antibody service** (*Reader Service No. 104*). The service includes the synthesis and purification (90–95 per cent purity) of a custom peptide, conjugation of the peptide antigen to a carrier protein, immunization of BALB-c mice, ELISA testing of the serum, fusion, primary and secondary hybridoma screening, cloning of two parent lines, characterization and final purification of the monoclonal antibody.

More than 2,000 **immunodiagnostic kits and reagents** for clinical and research laboratories are featured in the new 1992 catalogue from The Binding Site (*Reader Service No. 105*). Over 200 new products are listed, including autoimmune diagnostics, immunofixation kits and allergy diagnostics. A large range of affinity-purified antibodies and conjugates to mouse, rat and human immunoglobulins are also available, together with forensic-grade antispecies antibodies.

Sigma Immunochemicals presents its **1992 catalogue**, which lists over 1,300 antibodies, growth factors, enzyme substrate and buffer tablets, and related products (*Reader Service No. 106*). Over 100 new products are offered, including an expanded line of CD marker antibodies and the fluorochrome conjugates of these monoclonals. A complete line of mouse isotype controls is also offered for use in flow cytometry. New monoclonals for 1992 include antibodies to human proteins, cytoskeletal proteins, clotting factors, toxins and animal immunoglobulins. Multi-species antibody conjugates, antibody conjugates adsorbed with



Immunochemicals by mail order.

human and rat proteins, technical-grade immunoglobulins, and antibody purification kits are also available.

Pick up a copy of Rockland's latest **immunochemicals catalogue**, which features immunocytochemical reagents; IgG fractions, F(ab')<sub>2</sub> and affinity-purified antibodies — both unconjugated and conjugated to DTAF, horseradish peroxidase, biotin and alkaline phosphatase (*Reader Service No. 107*). Additional product listings include purified and immobilized proteins, antisera to enzymes and red blood cells; sterile serums, plasma and whole blood; tissue powders and much more.

## Mixedbag of monoclonals

BioGenex Laboratories has introduced a new monoclonal antibody that is designed for the **qualitative localization of P-105, a proliferation-associated nuclear antigen** expressed in proliferating cells and high-grade malignancies (*Reader Service No. 108*). It can be used with formalin-fixed, paraffin-embedded tissue or frozen tissue sections. Antibody to P-105 is directed against two polypeptides with molecular weights of 105 and 41 kD. This indicates that the antigen may occur in both a monomeric and dimeric form, or that the 41-kD protein is a proteolytic fragment of the 105-kD protein, BioGenex says. Anti-P-105 is known to stain preferentially interchromatin granules within the nuclear matrix, a region associated with RNA synthesis. The antigen, which may be involved in RNA metabolism or RNA transport, shows increased expression in proliferating cells and high-

grade malignancies. In malignant cells, increased expression of P-105 is not limited to late stages of the cell cycle, as it is in normal proliferating cells, but rather is found in most cells in all stages, the company notes. The antibody to P-105 should prove useful for identifying malignancies and for studying chromatin structure and malignant transformations. This mouse monoclonal is available in concentrated form or as a ready-to-use reagent that is optimized for use with the Super Sensitive biotin-streptavidin detection system.

SWant is a Swiss company that specializes in **antibodies directed against calcium-binding proteins** for use in immunohistochemistry and applications using immunoblots (*Reader Service No. 109*). The company is offering two highly potent and selective monoclonal antibodies against parvalbumin (mouse monoclonal 235, IgG1) and calbindin D-28k (mouse monoclonal 300, IgG1). These antibodies



Visualization of Purkinje cells of the cerebellum with SWant's monoclonal antibody against calbindin D-28k. (Note the labelling of the dendritic spines.)

have been shown to be excellent neuroanatomical markers, SWant says, and should be of interest to neuroanatomists. The advantage of these antibodies over those against amino acids and neuropeptides is that they stain the whole cytoplasm and even the thinnest ramifications (including dendritic spine and axon terminals), the company says. Both monoclonals are sold in 200- $\mu$ l quantities (lyophilized in ascites fluid) for \$100 (US).

Enzo has added a **monoclonal antibody that identifies neuron-specific enolase** (NSE) to its immunopathology product line (*Reader Service No. 110*). The antibody is directed against human gamma-gamma enolase and does not crossreact with either the alpha or beta subunits of the enzyme, Enzo says. The anti-NSE reacts with neurons, APUD cells, ganglion cells in the gastrointestinal tract, and both myelinated and unmyelinated nerve fibres. It should be a useful research tool for surgical pathologists in the identification of gangliomas, paragangliomas, oat-cell carcinomas and astrocytomas. In addition, it stains oligoastrocytomas, glioblastomas, meningiomas and

Schwanomas. The anti-NSE antibody can be used for staining formalin-fixed, paraffin-embedded tissue sections using standard detection methods such as Enzo's *ImmunoDETEK* staining kit for murine antibodies. The antibody is available as a ready-to-use reagent or in a concentrated format that can be titrated and then diluted to suit user requirements.

A new rat-specific **blood-brain barrier monoclonal antibody** is one of seven new reagents developed by researchers at



Rat cortex stained using the antibody to the blood-brain barrier antigen, which demonstrates many small blood vessels.

Sternberger Monoclonals, which are being distributed by Affiniti Research Products (*Reader Service No. 111*). The blood-brain barrier antibody can be used with frozen and de-paraffinized sections, and is specific for the endothelial protein found only in areas with blood-brain or blood-nerve barriers. Reactivity with this antibody disappears in lesions of experimental allergic encephalomyelitis, Affiniti says. Cocktails of monoclonal antibodies have been formulated to create the new PAN-NEURONAL and PAN-AXONAL MARKER reagents marketed by Affiniti. These reagents are suitable for use on frozen and de-paraffinized tissue sections and are particularly useful for anatomical tracing in mammalian and non-mammalian systems. The products should be of principal interest to anatomists, cardiovascular physiologists and neuroscientists who are interested in the functioning and diseases of the central and peripheral nervous system.

New for the new year is Zymed's **mouse monoclonal antibody to homing-cell adhesion molecule** (H-CAM or CD44) (*Reader Service No. 112*). The antibody reacts with human H-CAM and mouse GP85/Pgp-1. It inhibits H-CAM adhesion to hyaluronic acid and T-cell adhesion to erythrocytes. Zymed's antibody to homing-cell adhesion molecule may be useful in the detection of various lymphoid malignancies and metastatic cancer. The antibody is available purified or conjugated to fluorescein isothiocyanate, biotin or phycoerythrin.

Kamiya Biomedical has introduced a

new line of **monoclonal antibodies against cancer and cell-surface markers** (*Reader Service No. 113*). The line-up includes monoclonal antibodies against DUPAN-2, CEA, nonspecific crossreacting antigen, sialyl Le, SSEA-1, CALLA and other cancer and cell-surface markers. Kamiya says these monoclonals are available in a highly purified form and are useful research tools for cell staining, immunoprecipitation and other research applications.

ICN offers a new range of specific **monoclonal antibodies to rat cell-surface antigens and cellular adhesion molecules** (*Reader Service No. 114*). Anti-rat ICAM 1, a cell-surface ligand of the lymphocyte integrin LAF1, plays an important role in various cell-cell interactions of the immune system. Anti-rat LFA 1 alpha chain belongs to the family of leukocyte integrins and is known to be an important mediator of leukocyte physiology. Completing the line-up is anti-rat LFA 1 beta chain, a monoclonal antibody that also recognises MAC1, p150 and p95. ICN says the three monoclonals should be useful research tools in the study of disease states such as septic shock, cancer, organ rejection and chronic inflammation. All three are available in a purified form in 0.5-mg vials.

## Interleukins

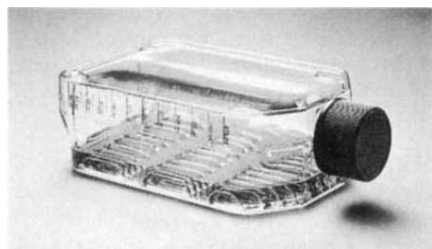
Human **interleukin-2** (IL-2), which is produced from pooled human blood leukocytes that have been tested and found to be negative for hepatitis-B surface antigen and antibodies to HIV-1, can now be purchased from Advanced Biotechnologies (*Reader Service No. 115*). The IL-2 is purified by a series of chromatographic procedures and is supplied in serum-free RPMI 1640, buffered with 25 mM HEPES. In this form, the IL-2 is appropriately glycosylated, has a biological activity that is standardized to internationally accepted BRMP units, and supports the long-term growth of human T cells in culture, the company says.

The Belgium company, Innogenetics, is offering **recombinant murine interleukin-6** (IL-6) produced in a baculovirus expression system (*Reader Service No. 116*). The mature protein murine IL-6 has a purity of greater than 95 per cent, the company says, and is purified by affinity chromatography using a monoclonal anti-IL-6 antibody coupled to Sepharose. The biological activity of recombinant murine IL-6, as determined in an inducing bioassay with 7TD1 cells, is  $10^6$  U mg<sup>-1</sup> recombinant protein. (One unit of activity is defined as the amount of material required to produce half maximal proliferation of the appropriate 7TD1 target cells.)



# Lab consumables

Cellon is distributing a new kind of **tissue culture flask** manufactured by In Vitro Scientific Products (IVSP) (*Reader Service No. 117*). The flask is blow-moulded from polyethylene terephthalate glycol (PETG), a flexible, shatter-resistant plastic. As blow-moulding forms each flask as a single piece, there are no leak-prone welded joints like those in two-piece, injection-moulded flasks, Cellon says. A proprietary surface treatment for PETG is designed to provide an improved growth surface that enhances cell growth and morphology for many cell types. Expanded-surface flasks (XPS) offer nearly twice the growth surface area of standard, flat-surface flasks, while still retaining original dimensions. Not only can use of the corrugated surface double incubator capacity and reduce handling time, but it may also produce higher proliferation rates for many cell lines, Cellon says. As with IVSP's XPS roller bottles, cells grow in monolayer over the corrugations and up the immersed side walls of the flask, as well as on the flat bottom surfaces. IVSP's flat-surface



XPS flasks from Cellon provide twice the growth surface area of conventional flat-surface flasks.

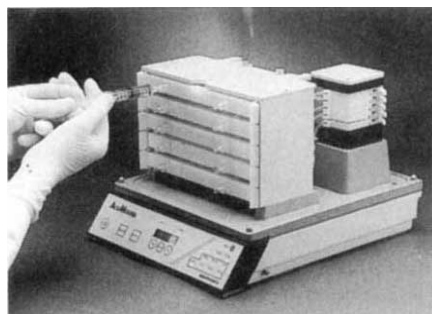
flasks are available in 75 cm<sup>2</sup> and 225 cm<sup>2</sup> sizes; expanded-surface flasks (XPS) of the same external dimensions provide 150 cm<sup>2</sup> and 450 cm<sup>2</sup> growth areas, respectively. The top panel can be cut away for easy access to the interior for harvesting purposes and to allow the user to perform *in situ* hybridizations.

Residues in your immunoplates should be a thing of the past with Nunc's new line of C-It **immunoplates** (*Reader Service No. 118*). With the redesign, the bottom

of the wells now have rounded corners but still retain a flat area in the middle. This, Nunc says, allows the user to improve the washing efficiency of the plate. The diameter of the wells have also been altered. The subsequent increase in reading pathway produces a higher signal and, thus, makes it possible to cut down on the amount of reagents used, the company claims.

# Gadgets and gear

The **production of monoclonal antibodies** need not be a chore, says Endotronics, when using the company's ACUMOUSE system, which is designed to provide an economical alternative to ascites (*Reader Service No. 119*). At the heart of the system is the new OXYCELL bioreactor



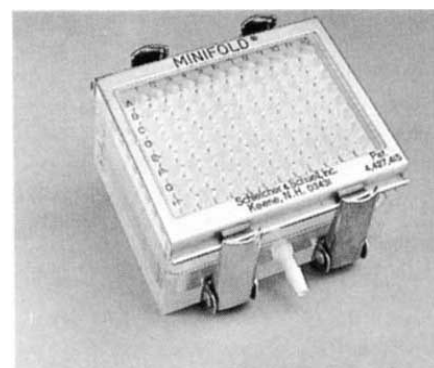
Endotronic's ascites alternative.

cassette, which delivers direct oxygenation to the cells. The system can accommodate up to five bioreactor cassettes for multiple cell-line or media screening.

Multisoft and Bichromatic are the two latest additions to Labsystems' Multiskan range of **microplate readers** (*Reader Service No. 120*). Multisoft, the successor to MCC/340, features two new developments: cartridge programmability and an agglutination scanning mode. Credit-card style programme cards permit sophisticated data reduction routines, including linear and cubic spine curve fits. When the agglutination mode is selected, an improved optical system and plate transport mechanism allows 20 readings across each well by utilising an innovative screening system that automatically stops the light beam size down to 1-mm diameter. Multisoft has a spectral range of

340 to 750 nm with enzyme immunoassay, kinetic and agglutination measuring modes as standard. Bichromatic, which supersedes the Multiskan Plus microplate reader, includes a filter wheel with seven filters, on-board mixing and a full kinetic capability as standard.

The Minifold I **96-well manifold** from Schleicher and Schuell is designed for use in nucleic acid dot-blots, hybridoma screening, TCA precipitations and particle entrapment (*Reader Service No. 121*). Use of the Minifold ensures the formation of discrete, uniform dot-blots carried out on filtration media such as S & S's NC nitrocellulose membranes, the company says. The O-ring design forms distinct, sealed circles with no lateral sample migration. When used in conjunction with the optional incubation plate, both filtration and incubation can be carried out in the Minifold device. The device is designed using the 96-well, microtitration (A-H, 1-12) orientation. Use of the Minifold device for ELISA-type immunoassays offers additional benefits over polystyrene-based ELISA tests, S & S says. ELISAs performed using S & S's BA-S NC nitrocellulose and the Minifold provide a permanent, flat



S & S's Minifold I device.

record of the test. Standard acrylic and autoclavable models are available. □

*These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal.*

# Synthetic Oligonucleotides

Custom made probes and primers, synthesised, purified and shipped in 3 days

**£3.50 per base**

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, blotin or digoxigenin labelling, data base analysis and full design facilities

Molecular Medicine Unit,  
King's College School of Medicine & Dentistry,  
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No. 56

# ADVERTISEMENTS

**PEPTIDE**  
PRODUCTS LTD.  
Custom Synthesis & Specialist  
Peptide Chemistry

# High quality standard services

Custom peptide synthesis  
HF cleavage and purification  
Peptide conjugation and antibody production  
Synthesis of modified amino acids and peptides

Peptide Products Ltd.  
David Hart Business Centre,  
Unit 2, Salisbury Road,  
Downton, SP5 3HT  
Wiltshire, England (U.K.)  
Tel: 0725 23021  
Fax: 0725 23024

Reader Service No. 19

**\$3.50/Base**

**Custom DNA  
Synthesis**

No Set-up Fee,  
Shipped in 24 hours.

**1-800-788-4362**

**1-800-786-4362 Fax**

**KEYSTONE**  
Laboratories, Inc.

Reader Service No. 33

NATURE · VOL 355 · 30 JANUARY 1992